Universidad de San Andrés

Departamento de Economía

Doctorado en Economía

Multidimensional Aspects of Welfare

Argentina 1991-2014

Autor: María Edo

DNI/Pas: 28031042

Director de Tesis: Walter Sosa Escudero

Victoria, Buenos Aires, Argentina

Noviembre 2016

Universidad de San Andrés

Doctorado en Economía

Multidimensional Aspects of Welfare

Argentina 1991-2014

María Edo

Advisor: Walter Sosa Escudero

Victoria, Buenos Aires, Argentina
Noviembre 2016

Contents

Acknowledgements

In the first place, I would like to thank my advisor, Walter Sosa Escudero. For his thoughtful guidance in terms of research, but also for initiating me in the art of teaching and for so many other generous academic advice. Without his appreciation of my heterodox background, traveling this road would not have been possible.

I am also extremely grateful to both Mariana Marchionni and Marcela Svarc. Their detailed guidance, advice and encouragement were crucial during the final stages of my PhD research. Their support was enriched by long and enlightening conversations about life.

The *Universidad de San Andrés* provided a motivating atmosphere to pursue my research. In particular, I would like to thank all of the professors at the Economics Department: Federico Weinschelbaum, Mariano Tommasi, Martín Rossi, Enrique Kawamura, Christian Ruzzier and Juan Carlos Hallak. Each of them in their own way helped me along this path. At the AAEP XLVII annual meeting, the 5th ECINEQ meeting as well as the 2016 RIDGE May Forum I received valuable comments from Leonardo Gasparini, María Ana Lugo and Verónica Frisancho. I would also like to thank Ricardo Crespo who offered me wise advice in the very early stages of my PhD. Thanks to MINCyT, CONICET and the Economics Department at *Universidad de San Andrés* for providing the financial support that made my research feasible.

A special thanks goes to my family. To Alejandro, my father, who helped me keep my inner compass always correctly calibrated. And to Nora, my mother, for being an outstanding model of professional and mother. I am sure she is enjoying this achievement.

Finally, I would like to express my most special gratitude to my husband Ricardo for his love, his patience, and his unfailing support in this and in every other commitment I have ever undertaken. And to my sons: Juan and Tomás who started this journey with me from the beginning and Santiago and Mateo that joined later on. Preparing customes for your amazing kindergarten shows, collecting useless plastic caps for your fabulous art projects, participating in thrilling science fairs and attending exciting soccer matches have paired with some of the most decisive deadlines, presentations and deliveries in my academic path. I am convinced that because of that, and not in spite of it, this long journey was possible.

Introduction

International experience shows that welfare is characterized by relative stability: changes may only be appreciated by assessing long periods of time. Argentina is a clear exception to this rule. In the last thirty years, its economy has traveled along a bumpy road: the debt crisis in the 80's, the hyperinflation episode in 1989, the 'convertibility plan' during the 1990s along with privatization of state-owned enterprises, the 2001 crisis and the following recovery and finally stagnation as from 2008. The impact of this sequence of crisis, relative stability and drastic political economy changes on different welfare aspects have been certainly very relevant, especially as from 1990.

As expected, this has awakened the interest of the academic community, giving rise to a vast literature. The unique experience of a country where income distribution could be seen in motion certainly provided an intellectual incentive. Studies aimed, primarily, at establishing the impacts of these changes in society as well as to study its causes.

Most of this literature has focused on issues of poverty and inequality, as a result of their sustained increase during the '90s. Without questioning the priority given to these issues, the rela and tive absence of other aspects of similar rank in the research agenda is remarkable. While poverty and inequality are two undeniably relevant phenomena of the income distribution, there are others whose study certainly deserves attention in terms of allowing for a more deep characterization of welfare changes in Argentina.

The aim of the present work is to make a contribution to this literature through the empirical study of two different yet interrelated aspects of well-being in Argentina that have received scarce attention. On the one hand, Chapter I and II will focus on the Argentinean middle class, assessing how this particular socio-economic group has performed during the last twenty years. This is a crucial phenomenon in a country like Argentina, where the relative size of the middle class has been a distinctive feature, at least in regional terms. On the other hand, Chapter III will concentrate on a different yet closely related issue: understanding the increase in secondary school attendance rates as from 2004. Motivation for this research clearly stems from the incidence of education on social and economical mobility, a key aspect of well-being closely related to the dynamics of the middle classes. Furthermore, it allows for a partial evaluation of the most relevant social policy implemented in the last 20 years in Argentina: the *Asignación Universal por Hijo*, a massive cash transfers program aimed at vulnerable children.

The first two chapters deal extensively with the identification of the middle class, a conceptual and methodological issue that is far from having achieved any agreement in the literature.

Chapter I, titled *Tracking the evolution of the middle class in Argentina 1991-2012*, concentrates on a unidimensional characterization based on income. Recognizing the importance of existing sociological studies in relation to the middle class, it pursues a more precise objective: to analyze the performance of a package of middle class measures based on income that have been proposed in the literature in recent decades, applied to the Argentinean case. Indeed, even reducing the definition of the middle class to a single attribute (income), the variety of proposed measures is not minor. This research, thus, seeks to show whether these differences are reflected in an empirical analysis of Argentina's middle class between 1991 and 2012, using three different indicators (the size of the middle class, its average income and the proportion of income it holds).

Chapter II, in contrast, follows a multidimensional path: *Measuring the Middle Class in many dimensions: Argentina 2004-2014*. This study was strongly motivated by the difficulties found in defining the middle class based solely on income. In fact, the poverty literature has already established the idea that only multidimensional measures may accurately reveal well-being. This becomes even more relevant when studying the middle class, since its unidimensional identification limits the analysis in at least two additional ways. On the one hand, even though certain agreement may be reached on establishing a lower bound in terms of income, setting an upper threshold is far less obvious. On the other hand, this implies disavowing the large and rich literature coming from the sociological and political theory realms that point to other dimensions as key in defining the middle class. Therefore, Chapter II contributes to the literature by presenting a new multidimensional approach to identify the middle class. By way of establishing multivariate α-quantiles based on a growth direction of increasing well-being we are able to multidimensionally identify the Argentinean middle class. We do so by relying on 19 variables associated to different aspects of welfare: per capita family income; sources of income, property and wealth; employment and education; dwelling characteristics and having a domestic employee. Chapter II also provides insight on the dimensionality of welfare: which and how many variables are relevant to capture welfare and, moreover, to identify the middle class? To this end a variable selection exercise is proposed in order to identify the smallest subset of dimensions that are able to reproduce results in terms of welfare and allocation to groups (poor, middle and upper class).

Finally, Chapter III: *Compulsory Education Laws or incentives from CCT programs? Explaining the rise in secondary school attendance in Argentina* sheds light on a totally different matter: the increase in secondary school attendance rates. Indeed, during the last 10 years a sizeable improvement in net attendance rates for children aged 15 through 17 took place. We show that even though in 2006 the National Education Law was passed making upper-

secondary education compulsory it is actually the *Asignación Universal por Hijo* (Universal Child Allowance, AUH) that may be held responsible for this improvement. The AUH is a massive conditional cash transfers program (CCT) targeted at children under 18 years old living in poor families whose parents are not registered workers in the formal employment sector. The program has no precedents in terms of coverage and relevance of the subsidies awarded. It currently benefits 29% of all children in the country, representing 0.8% of the country´s GDP. As any typical CCT program, the reception of the transfer is conditional on complying with children's health requirements and school attendance at all compulsory levels. The high economic incentives introduced by the AUH and its conditionalities may reduce the probability of dropping out of secondary school compared to the counterfactual situation in absence of the program. Using a difference-in-difference strategy we estimate the program effect on the probability of attending secondary school among eligible children aged 15 through 17.

Although this last Chapter certainly differs from the previous two both in focus and methodology, the subject addressed is certainly relevant to the core of the present thesis which is defined precisely by the study of multidimensional aspects of welfare in Argentina during the last two decades. It was motivated by the great void in terms of evaluation of such a large social program as the AUH. Even though some exploration of its effect on distributional aspects has been advanced, no rigorous attempt has been made so far to study its impact on other dimensions, such as the educational realm.

CHAPTER I. Tracking the evolution of the Middle Class in Argentina 1991-2012*

Abstract

In spite of its social and political relevance, economic studies of the middle class are scarce, mostly due to the conceptual and methodological difficulties in agreeing on a notion that allows empirical researchers to measure it, make valid comparisons or track its temporal evolution. Using data from Argentina, this study aims at assessing the performance of several practical notions of income based middle class. The dramatic changes suffered by the income distribution of this country during the last twenty years provide a suitable setting for this kind of analysis.

JEL classification: D3, I3, D6
Keywords: middle class, distribution, Argentina

* This article is joint work with Walter Sosa Escudero (UdeSA, CEDLAS-Universidad de La Plata and CONICET).

1. Introduction

The middle class as a group has occupied a relevant place in the political science and sociology literature during the last centuries. In contrast, only in the last few years have middle class studies gained relevance in the economic academic research agenda. Recent studies like Ravallion (2010), Atkinson and Brandolini (2011) and Banerjee and Duflo (2007) are indicative of the fact that the traditional focus on poverty-related issues and income inequality is shifting to this specific group. Furthermore, in Latin America the issue gained increasing visibility in 2012 with the publication of the Flagship Report of the World Bank "Economic Mobility and the rise of the Latin American Middle Class".

This renewed attention derives fundamentally from the key role assigned to the middle class in contemporaneous societies. Indeed, much is expected from this group. Some authors go as far as claiming that they represent the foundation on which democracy and market economy may flourish (Birdsall *et al.,* 2000). Others point to its capacity in terms of diminishing potential sources of conflict and polarization (Gigliarano and Muliere, 2012), as well as their central role in motorizing the economy through entrepreneurship and consumption (Banerjee and Duflo, 2007). The issue grew particular attention during the 1980s and early 1990s associated to the so-called middle-class decline that was claimed to occur in the US and other developed countries.

In spite of this general agreement on the relevance of middle classes in the political, social and economic processes, the economic academic community is far from having reached a minimum consensus on some conceptual and operational definition. Even if restricting the analysis to the income realm, huge theoretical discrepancies abound in defining the measurement of this segment of the income distribution. This lack of agreement seems to be driven by the same conceptual concerns that affect the empirical literature on the measurement of poverty, even for the simplified case of income based poverty. These concerns are exacerbated for the case of the middle class, since its operational definition now requires agreeing on an *upper bound*, above which lie the rich, or those rich enough to be considered out of the middle class.

Three major groups of definitions of income-based middle classes have been advanced in the economic literature: relative, absolute and hybrid measures. Under relative measures we group definitions coming from different theoretical backgrounds but which share one common feature: both thresholds move with changes in the income distribution. This group includes central tendency measures, as well as polarization and more standard relative measures. The second group mirrors the poverty literature, extending absolute definitions of poverty lines to measuring this portion of the income distribution. Finally, the third group includes measures that combine absolute and relative thresholds.

All these alternative conceptualizations of the middle class do not lack from theoretical interest, and are sustained by conceptual arguments worthy of attention. There is, however, an empirical question to address. Despite all these theoretical distinctions, do the different groups of measures yield different results when applied to empirical data? And, if so, what is driving these differences? Which measures should be preferred, and why? Finally, once this discussion is settled a new question arises: have the changes in the middle class welfare been paralleled by the predicted results in terms of polarization?

These will be the guiding questions in the present study. To answer them, we will focus on the temporal evolution of the middle class in one particular country. This is crucial to our analysis, given the large number of definitions and welfare indicators involved. Only by taking a deep look at just one country we will be able to disentangle the effects of adopting different conceptualizations of the middle class. In doing so, we consciously refrain from making country comparisons, which would certainly benefit the analysis. Further research should point in this direction.

The scenario chosen for the analysis is Argentina. This case is relevant not only because middle class issues are especially sensitive in this country[1] but most importantly for our purposes because it provides an exceptional setting for any income distribution analysis. Argentina diverges from the relative stability that characterizes the income distribution across countries, in which changes in this sense can only be evaluated by assessing significantly long periods. This country's moving distribution allows for a careful analysis of changes of the middle class in the last decades.

The analysis performed aims at contributing to the existent literature in two regards. On the one hand, it offers a rigorous empirical analysis of the consequences of adopting different income-based measures of the middle class. On the other hand, the methodological conclusions of this analysis allow for the characterization of the evolution of the middle class in Argentina in the last decades.

The rest of the article is organized in the following way: the next section describes the welfare indicator the measures of the middle class and the aggregation indicators chosen. Section 3 focuses on the dataset while section 4 presents the main empirical results. The last section concludes and point towards future research.

[1] Argentina remained an exception to the rest of Latin America by growing a large middle class during the second half of the 20th century. Nevertheless, it is claimed to have shrunk during recent decades. See New York Times, 11/12/2011, "What Latin America can teach us" and *La Nación*, 04/07/2004 *"Crisis y Clase Media: cuando poco queda en pie"*.

2. Measuring the Middle Class

Measuring the middle class is not an easy task since there is not an obvious way to define and quantify this group. In the sociological realm it has been characterized through the occupational structure. The definition of class has been more related to the ownership of assets and the type of occupation derived from it[2]. The economic approach, however, has taken an alternative route in welfare measurement. In particular, the poverty measurement literature has agreed on a procedure for measurement that involves the following steps: *(i)* the definition of a well-being indicator; *(ii)* the identification of the group of interest among the total population; *(iii)* the construction of an index of the relevant group using the available information[3]. Middle class measurement mirrors many of the difficulties that the poverty literature has dealt with. Thus, although recognizing the fundamental importance of the sociological tradition in these issues, we will concentrate on following the procedures upon which the economic literature has agreed:

i. Welfare indicator. In light of our purposes, the first of the exposed problems will not be explored in detail. We will base our analysis on a simple and reproducible income measure. Even though the copious literature on multidimensional welfare points towards the limitations of income to characterize well-being, income is still vastly used as a practical proxy to quantify welfare and its related notions like poverty or inequality. In light of the alarming scarcity of studies on the middle class, we will focus on income to concentrate the analysis on the comparison of alternative measures of the middle class. Extending the comparison to multivariate conceptions will be the focus of Chapter II.

ii. Identification. In terms of the identification of the middle class, a number of different economic notions have been advanced. Cruces *et al.* (2010) synthesize the different criteria adopted such as they are used in the applied literature and following the seminal classification proposed by Foster and Wolfson (2009). Broadly speaking, middle class economic definitions can be grouped into three categories: relative, absolute and hybrid measures. Relative measures were the first ones applied in income-based middle class studies. Central tendency (CT) measures have been the pioneering criteria, originally designed to assess the claimed decline of the middle class in the US and other developed countries as from the 80's. These definitions classified as middle class those individuals whose income fell within some range involving the median, being typically symmetrical around it. Although the lower and upper bound have been set in several ways in different contexts, the literature seems to have converged to a 75% to 125% interval around the median, as proposed by an influential early paper by Thurrow (1987). This discussion on the presumed disappearance of the middle class in the USA and Canada

[2] For instance, see Giddens (1981), Erikson and Goldthorpe (2002).
[3] The identification of the latter two problems in the poverty literature is due to Sen (1976).

during the 80's also gave place to the emergence of measures of this group derived from the polarization literature. Foster and Wolfson (1992) linked the polarization of the income distribution within a society to the disappearance of its middle classes. Later on the concept of polarization has been further expanded with the measures developed by Esteban and Ray (1994) and Esteban, Gradin and Ray (1999)[4]. They provide a theory of polarization measurement, distinguishing it from inequality measurement. The polarization measure is axiomatically derived from the identification of groups that are distant from each other (alienation) but that have a similar internal composition (identification)[5]. Polarization is defined as some increasing function of these two features. Cruces *et al*. (2009) use one particular case of these measures: departing from exogenously establishing the number of groups within a society to be three, the measure endogenously determines the cut-off points that define them. The resulting groups are the ones with the highest identification within them and highest possible alienation among them. Given the fact that the number of groups was set to three, their cut-off points are used by Cruces *et al*. (2009) to derive the lower and upper thresholds of the middle class. Several relative measures have been proposed in the literature as well, based on quantiles of the income distribution, typically leaving out the lowest and highest deciles or quintiles. The idea guiding this kind of partition is that the first two deciles clearly belong to the most deprived while the upper two deciles are capturing the long tail of the income distribution. Barro and Easterly (2001) propose classifying as middle class those who lie between the 3rd and 8th decile of the distribution. Although central tendency, polarization and properly called relative measures differ greatly in their theoretical backgrounds, they share one common feature that is crucial in terms of middle class measurement: all of them rely on implicit poverty and richness lines that move in accordance with the income distribution. Section 4 will provide some insight on the importance of this feature.

More recently, the focus on middle class studies has followed the developments made in the poverty measurement field. Ravallion (2009) proposes the extension of absolute poverty lines, setting u$2.5 (PPP) as the lower bound and u$13 (PPP) as the upper bound, pointing to the risk of applying definitions suitable for rich countries (i.e., central tendency measures) to those in the developing world. The lower bound builds on the theoretical development of absolute poverty lines: it is based on nutritional requirements and non-food needs, reflecting the median of poverty lines in 70 developing countries. The upper bound is the US poverty line, which

[4] The theory of polarization measurement was developed further still in a subsequent paper by Esteban, Duclos and Ray (2004) for the case in which asset distributions can be described or estimated (parametrically or otherwise). They also provide statistical inference results that can be used to compare indices of polarization across societies, time or economic regimes.

[5] Other endogenous measures based on cluster and principal components analysis as well as subjective and multidimensional measures of the middle class have been advanced but are not considered in this work in light of the aim at hand.

ensures that no one considered "poor" in the US could be classified as rich in developing countries. Banerjee and Duflo (2007) establish a similar measure, ranging from u$2.5 (PPP) to u$10 (PPP)[6]. Quite recently, and following studies focused on vulnerability to poverty (López-Calva and Ortiz-Juarez, 2011), new (higher) thresholds have been proposed. For instance, Loayza *et al*. (2012) set the lower threshold at u$10 a day, following the idea that the behavior of those individuals lying close to the poverty line, whether classified as poor or middle class, may not differ. The authors avoid establishing a richness line arguing that it might lead to artificially small middle classes in rich countries. Instead, the regional Flagship Report of the World Bank for Latin American and the Caribbean define U$50 (PPP) a day as the upper middle class threshold, while keeping the U$10 a day as the poverty line.

Table 1. Middle Class Measures

MIDDLE CLASS MEASURES					
Type	**Definition**	**Thresholds**		**References**	**Accronyms**
		Lower	**Upper**		
Relative	Based on measures of central tendency.	75% of median income	125% of median income	Thurrow (1987)	CT
	Based on polarization measures.	Lower cut-off point of polarization measure.	Higher cut-off point of polarization measure.	Cruces et al. (2010)	PLZ
	Based on quantiles of the income distribution.	3rd decile	8th decile	Barro (1999); Easterly (2001)	BE
Absolute	Based on absolute thresholds.	2.5 USD (PPP)	10 USD (PPP)	Banerjee and Duflo (2007)	BD
		2.5 USD (PPP)	13 USD (PPP)	Ravallion (2009)	RA
		10 USD (PPP)	50 USD (PPP)	LAC Flagship Report (The World Bank)	FL
Mixed Strategies	Based on absolute and relative thresholds.	Official poverty line	90^{th} percentile	Sosa Escudero and Petralia (2010)	SP
		10 USD (PPP)	95^{th} percentile	Birdsall (2010)	BIR

The latter group follows mixed strategies as a way to deal with the difficulties derived from establishing a richness line. For instance, Sosa Escudero and Petralia (2010) propose the use of the national poverty line as a bottom bound and the 90^{th} percentile as the upper threshold. The middle class are trivially the mirror of the poor leaving out those included in the farthest right

[6] The authors actually define two alternative segments as well, U$4-U$10 and U$6-U$10.

part of the distribution. In line with this hybrid definition of the middle class, Birdsall (2010) establishes u$10 a day as the lower bound on the grounds that, even though admittedly ad-hoc, it provides the minimum income to ensure economic security and thus leaves space for the individual to exercise political rights. The upper limit is set at the 95th percentile arguing that in this way the portion of the population whose income most likely comes from sources other than productive labor (such as inherited wealth, past or current economic rents, etc.) is excluded from the middle class.

iii. Aggregation. In order to assess the different middle class measures proposed in the literature, it isrequired to build an index that allows for their comparison. In this study we will focus in three different criteria for assessing the well-being of the middle class across time: its *size*, its *mean income* and its *share of income*. The *size* of the middle class refers to the proportion of individuals classified as belonging to the "middle class"[7] according to each of the measures under analysis. The *mean income* reflects the well-being of this group across time. The *income share* sheds light on the standing of the middle class regarding the other groups (i.e., the "poor" and the "upper class").

3. The data

To track the temporal evolution of middle class measures in Argentina micro data coming from the *Encuesta Permanente de Hogares* (EPH) will be used. The survey covers labor as well as self-employment income and monetary transfers. In particular, the analysis will be based on data for Greater Buenos Aires (GBA), rather than for the country as a whole. Questioning regarding the representative nature of the data may rapidly arise. Gasparini and Cruces (2009), however, show that the evolution of aggregate distribution figures remains virtually unaltered to the inclusion of all regions. The income definition used throughout this analysis is per capita income, which has been deflated for comparability[8]. In order to keep the analysis as simple as possible and to make it easily reproducible, no further adjustments have been made.

4. Empirical results

The central question guiding this paper is whether the theoretical differences of the alternative types of middle class measures are reflected empirically. That is, we will follow the temporal

[7] This would be an extension of what is known in the poverty literature as "headcount index".

[8] Income has been deflated by the official price consumer index for the range 1991-2007. During the period going from that year to 2012 the reliability of the data released by the national statistics office (INDEC) has been seriously questioned. For those years, we follow data provided by the *Center for Distributive, Labor and Social Studies (CEDLAS)*, through their *Socio-Economic Database for Latin America and the Caribbean (SEDLAC)* developed in partnership with the World Bank. For further information: www.cedlas. econo.unlp.edu.ar

trends of the middle class in Argentina, as defined by each measure, and evaluate whether they differ or not in the three indicators chosen: its size, its mean income and its income share.

Before going any further, it is important to clarify what we mean by "empirical differences". Alternative measures can differ across time in two dimensions: their levels and their trends. Hence, it might be the case that though suggesting markedly different levels of the size of the middle class for a certain region, the evolution depicted by these alternative measures coincide. In one extreme, all measures of the middle class might differ in their levels but reveal the same temporal behavior, so in spite of being unable to provide consistent measures of the level (size or performance) of the middle class for a particular point in time, they are useful to quantify its evolution. On the other extreme case, each measure suggests different levels and patterns, being a reflection of alternative conceptualizations of the importance and evolution of the middle class. A similar concern holds for the problem of measuring inequality[9]. Given the lack of agreement among middle class measures, it seems natural to explore the empirical differences of their *temporal evolutions*, rather than differences in their levels. Thus, the central aim of this paper will be to understand whether the alternative definitions of the middle class convey marked differences in the *pattern* described across time.

Figure 1 shows the temporal evolution of the middle class in Argentina between 1991 and 2012 as measured by the different notions described above. Each column represents a particular group of measures: relative measures, absolute measures and hybrid respectively. Rows represent different welfare indexes: size of the middle class, its mean income and its mean income share. Note that in addition to all measures described in section 2, Figure 1 includes an alternative definition of the middle class: the "mid 50%". That is, the middle class is plainly defined as the group that represents the middle 50% of the income distribution with the aim of allowing for a baseline comparison in all groups of measures.

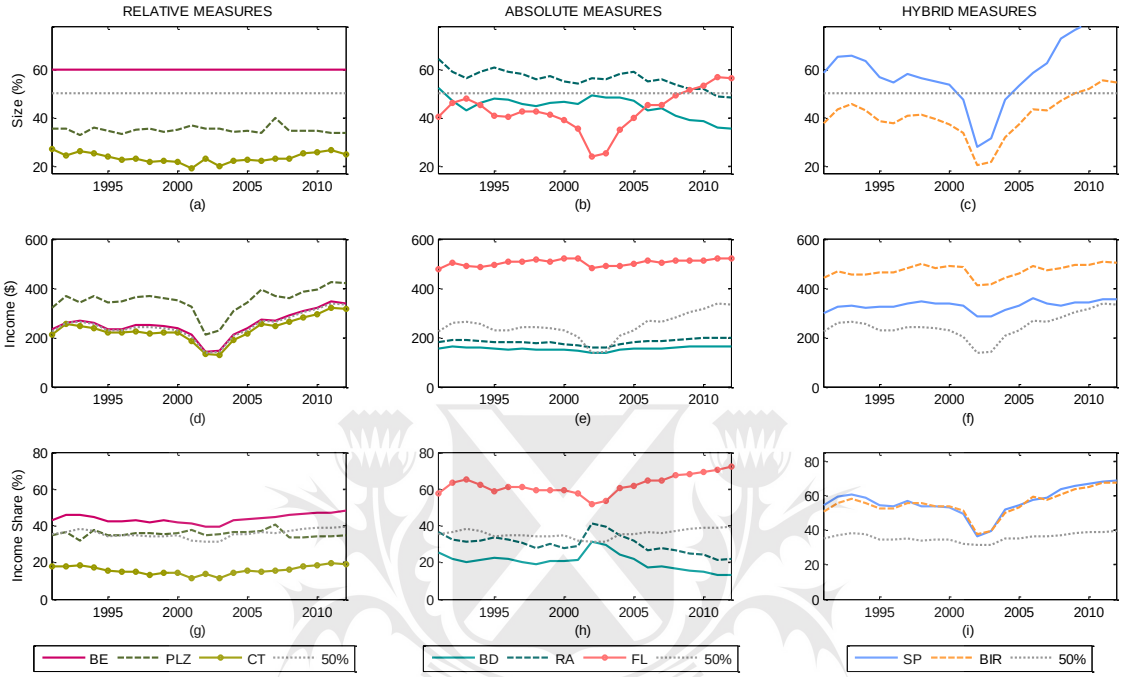
A rather unclear picture of the changes fared by the middle class in Argentina in the last decades emerges. Indeed, the empirical results are far from being consistent across the different indicators and theoretical groups of measures. Even within the same groups sharp divergences can be found.

Relative measures show similar patterns among them, though levels vary greatly. If the size and the income share of the middle class is considered, according to these measures the middle class

[9] For example, Gasparini and Sosa Escudero (2001) show that for the case of Argentina, alternative measures of inequality suggest different estimated levels but same temporal evolution between 1991 and 1994, but lead to alternative perceptions of this evolution during the period 1994-1998.

went through quite a stable period[10]. Nevertheless, the mean income of this group seems to have suffered a strong setback during the 2001-2 crisis, recovering and improving its level by 2012.

Figure 1. Middle Class In Argentina, 1991-2012



Source: own calculations based on EPH.

Notes: each column represents a category of Middle Class (MC) measures (Relative, Absolute, Hybrid) while each row represents a different indicator (Size, Mean Income, Income Share). *(a)* Size of MC according to Relative Measures; *(b)* Size of MC according to Absolute Measures; *(c)* Size of MC according to Hybrid measures; *(d)* Income of MC according to Relative Measures; *(e)* Income of MC according to Absolute Measures; *(f)* Income of MC according to Hybrid Measures; *(g)* Income share of MC according to Relative Measures; *(h)* Income share of MC according to Absolute Measures; *(i)* Income Share according to Hybrid measures. Se tables A.1, A.2 and A.3 in the Appendix for the complete series.

Absolute measures, instead, seem to disagree among them. When considering RA and BD measures, it may be concluded that the size and the income share suffered a persistent –yet soft- decay during the period while in terms of mean income it remained fairly unaltered. Conversely, the FL measure differs both in levels and in trend: not only is the middle class larger and better off in terms of income and income share, but also it seems to have improved its condition between 1991 and 2012 while suffering a decay-recovery pattern during the 2001-2 crisis.

[10] Although this conclusion is evidently trivial in the case of the BE measure for the size indicator (the measure is defined as the mid-60% of the population), this is not obvious in the case of the CT and PLZ measures. Furthermore, the relative stability of the income share indicator for the BE measure is not trivial either.

Hybrid measures differ in level while showing a very consistent picture not only among them but across indicators as well. For these measures, the middle class in Argentina was better off in 2012 as compared to 1991 in all dimensions (size, income and income share) but shows clear signs of having suffered the 2001-02 crisis.

A remarkable fact may be appreciated in these graphs. As rough as it may seem, defining the middle class as being "the middle 50%" may not lose much in front of more sophisticated definitions: the evolution of this group according to this definition is right in the middle of *all* the rest.

So far it is not possible to draw consistent conclusions regarding the path followed by the Argentinean middle class during the last two decades. Was the middle class in Argentina larger in 2012 as compared to 1991? Relative measures say 'no', hybrid measures say 'yes'. Did this group enjoy stability across the period? Not according to hybrid measures, but relative measures would disagree in terms of income share.

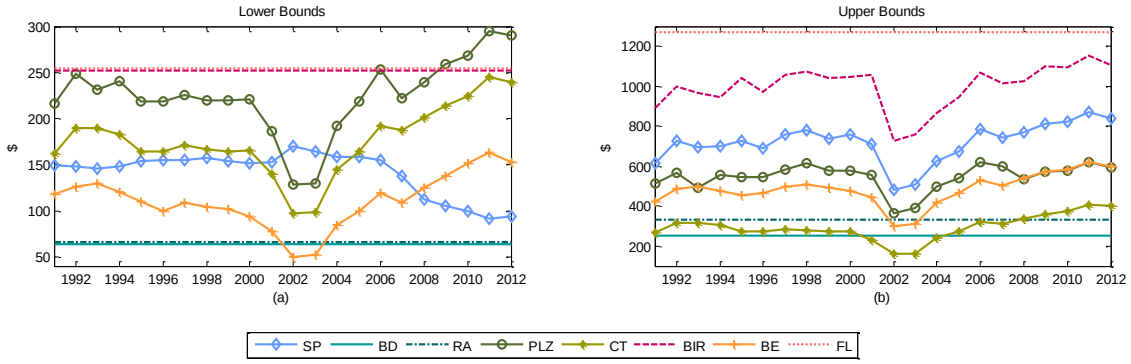
In front of this confusing panorama, some natural questions emerge: what is driving these -apparently- contradictory results? Are the different measures conveying opposite information? Or are they actually measuring different things? In any case, which of them is most suitable for assessing the course of the Argentinean middle class?

4.1 Reasons behind empirical differences

The first step necessary to fully grasp the dynamics underlying the empirical differences shown is to thoroughly understand the implications of choosing different upper and lower thresholds.

Figure 2 shows the poverty and richness lines for all the measures under consideration -whether implicit or explicit- expressed in argentine pesos (deflated to 2005). As expected, two types of lines can be observed: stable and more volatile thresholds. The first group includes absolute thresholds. RA, BD and FL rely on these thresholds both for the upper and the lower threshold. The SP and BIR measures rely on an absolute threshold only regarding the poverty line. The second group is made up of lines derived from the shape of the income distribution at each point in time. CT, PLZ and BE rely on this kind of thresholds in both limits while the SP and BIR only impose these type of lines in their upper limit. Note that this second group shows a strong decay-recovery pattern during the 2001-02 crisis. With a clear notion of what is going on 'behind the scenes' it is now easier to disentangle the inconsistencies found in the previous section.

Figure 2. Lower and Upper Bounds of Middle Class Measures in Argentina, 1991-2012



Source: own calculations based on EPH.

Notes: Panel (a) shows Lower Bounds for Middle Class Measures; Panel (b) Upper Bounds for Middle Class Measures.

Relative Measures

Figures *1(a)*, *1(d)* and *1(g)* show that even though coming from different theoretical backgrounds this group of measures show a similar pattern of the middle class among them while conveying a rather inconsistent picture of the middle class across indicators. Puzzling as it may seem, these outcomes are a natural derivation of the very definition of relative measures. The factor explaining both phenomena is the same: all three measures vary their upper and lower bounds with movements in the income distribution. It can be observed in Figure 2 that both the poverty and upper lines fall during the 2001-02 period.

This implies that relative measures will remain rather uninformative in terms of the path followed by the size and the income share of the middle class. Indeed, the poverty and richness lines are mimicking the shift of the income distribution to the left during the 2001-02 crisis, leaving both the size and the income share of the middle class virtually unaltered (see Figure A.1 included the Appendix). The extreme case for this source of stability of the size of the middle class is represented by the BE measure: by construction, the upper and lower thresholds mimic the movements of the income distribution so as to keep its size unaltered at 60%.

To sum up, the relative measures group remains rather uninformative in terms of size and income share of the middle class across time. Its very definition implies that these indicators will not change across time, showing a rather stable picture even in front of great crisis as the one suffered in Argentina in 2001-02. Therefore, when resorting to relative measures the only relevant indicator to be observed is the mean income of the middle class.

Absolute measures

Conversely, this group of measures comes from the same theoretical background and yet shows discrepancies not only in levels but also in the pattern followed by the middle class. The key to understanding the different empirical results is the level of the threshold chosen. In particular, note that the upper line set by the RA measure (which is very near the BD one) actually represents the poverty line for FL definition. This implies that even though based on the same theoretical notions, these measures are identifying different groups. In fact for the FL measure –almost- all individuals classified as middle class by RA and BD would be classified as poor.

At this point, therefore, it is crucial to establish which group are RA and BD really identifying in the case of Argentina. In this sense, Figure *2 (a)* clearly shows RA/BD lower threshold is the lowest of all measures across the whole period. The official poverty line set by the country is approximately three times higher than the RA/BD lower threshold. Therefore, these measures are largely capturing individuals regarded as poor by all other measures. Consequently, they are more probably reflecting changes in the welfare of the –not so- poor than that of the middle class. It seems then that Ravallion's cautionary note should be extended: it is clear that importing measures from rich countries may not be suitable for developing ones, but the converse is also true, that is, imposing definitions from far less developed countries may also lead to inconsistent conclusions.

The FL definition, instead, sets a poverty threshold more consistent with the middle class in Argentina. One word of caution remains nonetheless: establishing a lower and an upper absolute threshold implies that the indicators will suffer from composition effects. For instance, in terms of size opposite forces will come into play if the income distribution moves to the left: the middle class will shrink as a result of people falling into poverty but it will increase insofar as rich people become middle class. The final outcome will require empirical verification[11].

Hybrid measures

This latter group of measures is consistent in terms of trends and across indicators, and even quite close in terms of levels. Furthermore, these measures do not suffer from the problems exposed in the two previous groups. Unlike pure relative measures, its absolute poverty line ensures the utility of the size and income share indicators. Furthermore, its upper relative threshold ensures that these measures do not suffer from composition effects.

[11] In the FL case the composition effect tends to be weak since the poverty threshold is set in a very dense part of the income distribution while the upper limit is quite far to the right. This implies that movements of the income distribution to the left will have important effects regarding middle class individuals becoming poor but this will not be the case in the right tail of the distribution. For clarification, see Figures A.1 and A.2 in the appendix.

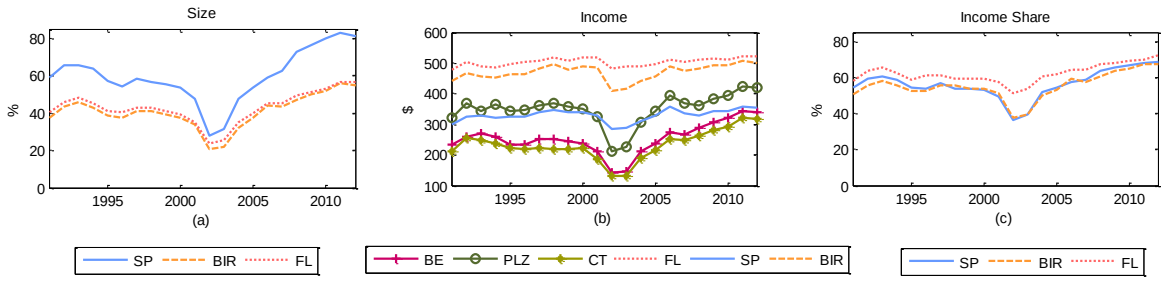
4.2 The course of the Argentinean Middle Class 1991-2012

The puzzle set by Figure 1 seems to become clearer now. The previous analysis reveals that not all the empirical results shown are informative of the course of the Argentinean middle class during the period under study. In particular, the size and income share indicators for relative measures are useless (Figures *1(a)* and *1(g)*) and RA/BD measures should not be considered as identifying the middle class for a country such as Argentina. When omitting the mentioned measures and indicators, the picture becomes rather consistent, as shown in Figure 3.

Once again, levels tend to differ –and by large- among the different measures, but the patterns followed are quite similar, both among measures and across indicators. In broad terms, it can be appreciated that the middle class seems to be better off in 2012 as compared to 1991. Furthermore, the middle class apparently did not enjoy this level of welfare at any point during the last 20 years, both in terms of size and income.

The path followed, however, was a rather bumpy road. During the period under analysis the Argentinean middle class suffered from a great instability in all dimensions. Three stages may be clearly distinguished: the 90's, the 2001-02 crisis and the 2003-2012 period. The first stage shows a rapid yet short improvement of the middle class welfare, most probably due to the recovery of the great hyperinflation crisis of 1989-90. From then on both the size and income of the middle class seems to start a rather soft downward road[12] which will finally lead to the tragic 2001-2 crisis. The middle class' welfare collapses during this second stage, showing recovery signs as from 2003. From that year onwards, the size and income of the middle class progressively improves reaching by 2012 its highest level in those two decades.

Figure 3. Middle Class in Argentina for Selected Measures, 1991-2012



Source: own calculations based on EPH.
Notes: (a) *Size* of MC according to selected measures; (b) *Income* of the MC according to selected measures; (c) *Income share* of the MC according to selected measures.

[12] FL and BIR measures in terms of mean income are the only measures/indicator that do not show this decay as from 1995.

4.3 Rising Middle Class, Rising Polarization?

So far it is clear that the middle class in Argentina has improved its well-being in many dimensions between 1991 and 2012. In particular, as from 2003 its size, mean income and income share have increased continuously showing by 2012 the highest levels in two decades. According to the literature, these improvements in the size and well-being of this group should be followed by higher levels of social cohesion (Barro, 1999; Torche and López-Calva, 2010) and lower levels of polarization (Wolfson, 1992; Cruces *et al*., 2011). In this section we explore whether this association holds true in the case of Argentina.

In defining polarization we follow the path outlined by Esteban and Ray (1994). As already stated, according to these authors polarization can be thought of as a measure that reflects simultaneously the distance among (alienation) and within (identification) groups at a particular time within a society. These two features work in opposite directions: larger distances among groups would raise polarization (through higher alienation) while increasing distances within groups would tend to reduce it (through lower identification). The authors propose several measures that aim to capture some sort of "clustering" among the income distribution.

Zhang and Kanbur (2001) develop an alternative measure of polarization for exogenously defined groupings, such as the 'inland-coastland' and 'urban-rural' divide in China. The measure proposed is the ratio of the between to within groups' inequality (where inequality is defined in terms of the Generalized Entropy Index): Note that while the authors do not make it explicit, the 'between group inequality' may be interpreted as the 'alienation' component of polarization in Esteban and Ray (1994) while the 'within group inequality' may be thought of as the identification component.

Given the nature of our analysis, we follow a similar approach to that of Zhang and Kanbur (2001). Indeed, we have exogenously defined –implicitly or explicity- three groups: the poor, the middle and the upper class. Thus, we intend to assess whether polarization has reduced among these groups. In particular, we attempt to answer several questions: has the distance among these groups remained constant across the period (alienation component)? If not, which groups have contributed more to its variation? Has the internal composition within each group changed (identification component)?

To answer these questions we resort to an ANOVA model which allows us to explore the variance between and within these three groups across the last two decades. The proposed model relates each individual's income to its group of belonging in each year[13]:

$$income_i = \beta_1 poorclass_i + \beta_2 middleclass_i + \beta_3 upperclass_i + \mu_i$$

Where *income* measures the deflated per capita family income; *poorclass, middleclass, upperclass* are dummies assigning 1 to the individual if he/she belongs to that group and 0 otherwise. The model is estimated separately for each year.

The model proposed may shed light in terms of polarization. In particular, we claim that the F test of such a model may be interpreted as a Polarization measure *à la* Zhang and Kanbur (1994), inasmuch as it as it assesses the relationship between the inter-groups inequality and the within-groups inequality:

$$Polarization\ Measure: F\ Statistic = \frac{between\ groups\ variation}{within\ groups\ variation} = \frac{MSModel}{MSResidual}$$

MSModel: SSM Mean Square | MSResidual: SSR Mean Square
SSM: Sum of Squares Model | SSR: Sum of Squares Residual

In a sense, the F test is comparing the 'alienation' between groups with their level of 'identification'. Indeed it would rise if the former rose and the latter remained unaltered (and vice-versa). If both forces are at work at the same time, the total effect will depend on the magnitude of each of them.

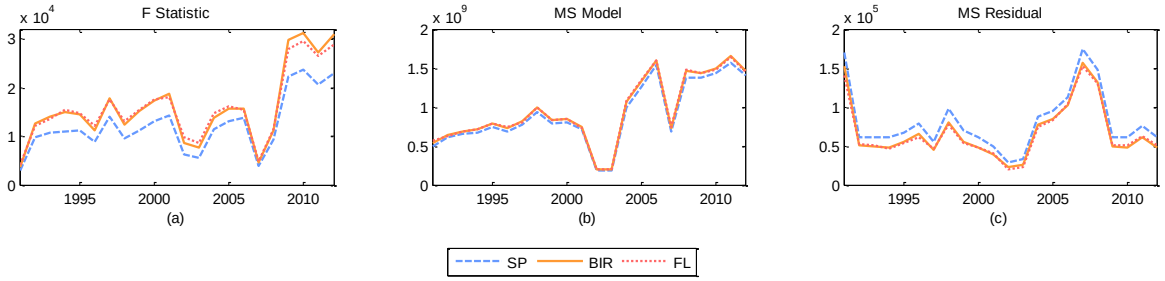
Furthermore, the model allows not only for assessing polarization at a more general level, but it can also point to which groups have been contributing to changes in the measure. For instance, the estimated coefficients are in fact the mean income of each group at every point in time. This provides insight on which pair of groups has become more close/more distant across time. Further still, the estimated standard errors for each coefficient may be thought of as the internal composition parameter (i.e., the 'identification') since they reflect the variance within each group.

The evidence for Argentina is presented in Figure 4, 5 and 6[14]. According to our measure (the F Statistic) polarization shows a rather stable –yet oscillating- pattern from 1991 to 2006 with its

[13] We estimated three alternative models obtaining similar results (taking the natural logarithm of income and models taking the Middle Class as the base category). The model used in the analysis was chosen on grounds of its straight-forward interpretation.
[14] In this section we only included the measures that performed best across all indicators: SP, BIR and FL.

lowest level in 2007, as shown in figure 4. From 2008 onwards, however, it increased continuously, reaching unprecedented levels. It is interesting to identify which forces are contributing to the rising polarization levels. Figure 4 show this is the result of unprecedented 'alienation' (i.e. distance between groups' mean incomes) combined with relative high levels of 'identification' (i.e., income inequality within each group).

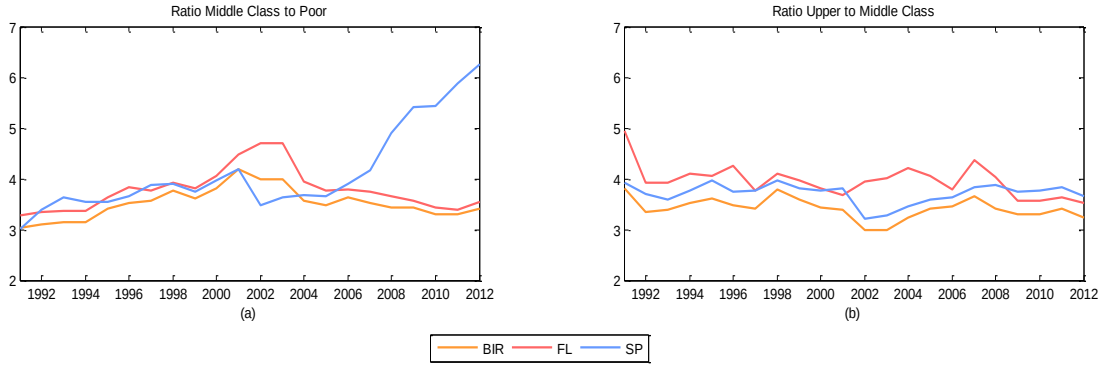
Figure 4. ANOVA Model



Source: own calculations based on EPH.
Notes: (a) F Statistic of the ANOVA Model N.1; (b) Mean Square of the Explained Sum of Squares of the Model; (c) Mean Square of the Residual Sum of Squares of the Model.

This naturally prompts questions regarding which groups are contributing to the distance among mean incomes as well as within which of them has the internal cohesion increased. As stated before, these questions may be answered by resorting to the estimated coefficients of our ANOVA model as well as to their standard errors. Figure 5 shows the mean income ratio between the Middle Class and the Poor and the Middle Class and the Upper Class respectively. Focusing on the 2008-2012 period it can be clearly seen that while the difference between the Upper and the Middle Class is falling and has reached lower levels than during the 90's the distance between the poor and the Middle Class did not revert to those lowest levels. Indeed, in 2012 the ratio was between 8% and 12% higher than in 1991 (its minimum level during the period) according to the BIR and FL measures respectively. The SP measure describes an even more dramatic scenario: the distance in 2012 doubles the one present in 1991.

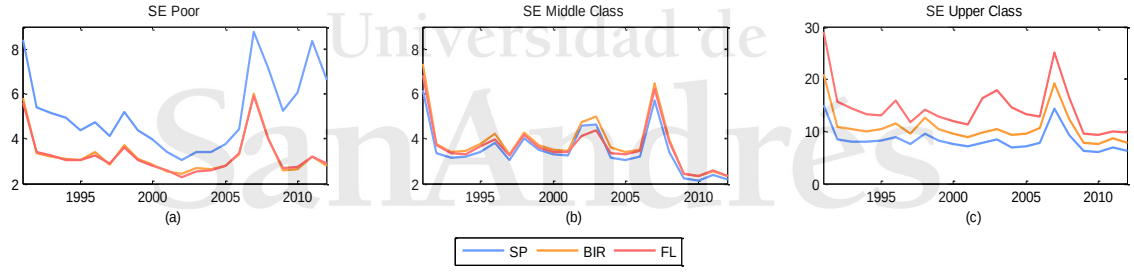
Figure 5. ANOVA Model – Mean Income Ratio



Source: own calculations based on EPH.
Notes: (a) Mean Income Ratio of Middle Class to poor; (b) Mean Income Ratio of Upper to Middle Class.

Regarding internal cohesion, Figure 6 shows that both the middle and the upper class show the lowest variance during 2008-2012. Even though the poor are also characterized by low levels of inequality within the group for this sub-period, it is clear that the middle and the upper class are generating the increase in the "identification" of the groups that had already been documented by Figure *4(c)*.

Figure 6. ANOVA Model – Standard Errors



Source: own calculations based on EPH.

Notes: (a) ANOVA Model Standard Errors for the Poor groop; (b) ANOVA Model Standard Errors for the Middle Class group; (c) ANOVA Model Standard Errors for the Upper Class group.

To sum up, the unprecedented levels in terms of size and economic performance reached by the middle class during 2008-2012 have not been followed by the larger economic homogeneity suggested in the literature. In Argentina at least, however, the improvements in all dimensions shown by the middle class as from 2008 have been paralleled by high levels of polarization. The distance among the three groups (poor, middle and upper class) has been increasing together with their internal cohesion. This combination of both higher alienation and identification may be described as an increase in economic polarization. In particular, it seems that the poor are the

ones being left behind in terms of their mean income while the upper and the middle class are the groups that have increased their internal cohesion the most.

5. Conclusion

The paper reviews eight different income-based definitions of the middle class such as existent in the economic literature with the aim of assessing their empirical differences. The analysis carried out using the case of Argentina shows that hybrid measures (i.e., those combining an absolute lower threshold with a relative upper one) are those that perform best across the three indicators chosen: size, mean income and income share of the middle class.

Regarding the course of the Argentinean middle class, the results show that while the 90's were characterized by a slow decay crowned by the collapse suffered during the crisis, from then on all indicators improved, reaching by 2012 unprecedented levels of welfare in the last 20 years. This, however, does not seem to have been paralleled by a reduction in polarization as suggested by the literature, at least not when analyzing the movements of the three groups defined: the poor, the middle and the upper class.

Further research should point in two directions. On the one hand, extending middle class measures into the multidimensional realm seems imperative given the added difficulties posed by the need of defining an upper threshold. In so doing, the economic approach to defining the middle class may benefit from the rich sociologic tradition on this regard. On the other hand, further exploration of the internal changes suffered by the Argentinean middle class as well as their link to polarization issues would certainly yield interesting insights of the social changes in Argentina over the last twenty years.

References

Atkinson, A. and Brandolini, A.: On the identification of the Middle Class. ECINEQ Working Paper No. 217 (2011)

Banerjee, A. and Duflo, E.: What is Middle Class About the Middle Classes Around the World? MIT Department of Economics Working Paper No. 07-29 (2007)

Barro, R.: Determinants of Democracy. Journal of Political Economy 107(S6), 158-183 (1999)

Birdsall, N.: The (Indispensable) Middle Class in Developing Countries; or, The Rich and the Rest, Not the Poor and the Rest. Center for Global Development Working Paper No. 207 (2010)

Birdsall, N., Graham, C. and Pettinato, S.: Stuck in the Tunnel: Is Globalization Muddling the Middle Class? Center on Social and Economic Dynamics Working Paper No. 14 (2000)

Blackburn, M. and Bloom, D.E.: What is Happening to the Middle Class? American Demographics 7(1), 19–25 (1985)

Callorda, F. and Caruso, G.: Does the Middle Class Really Exist? A Cluster-Analysis Approach. Universidad de San Andrés mimeo (2009)

Cruces, G., López-Calva, L. and Battiston, D.: What is Middle Class About the Middle Classes Around the World? UNDP / Research for Public Policy Inclusive Development Working Paper ID-03-2010 (2011)

Cruces, G., Pérez Truglia, R. and Tetaz, M.: Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. IZA Discussion Paper No. 5699 (2011)

Easterly, W.: The Middle Class Consensus and Economic Development. Journal of Economic Growth. 6(4), 317-35 (2001)

Esteban, J. and Ray, D.: On the Measurement of Polarization. Econometrica 62, 819-851 (1994)

Fajardo, J. and Lora, E.: Understanding the Latin American Middle Classes: Reality and Perception. Inter-American Development Bank mimeo (2010)

Ferreira F. H. G., Messina J., Rigolini J., López-Calva L. F., Lugo M. A., Vakis R.: Economic Mobility and the Rise of the Latin American Middle Class. Preliminary Findings LAC Regional Flagship 2012, The World Bank (2012)

Foster, J. and Wolfson, M.: Polarization and the decline of the Middle Class: Canada and the US. Journal of Economic Inequality 8(2), 247-273 (2009)

Gasparini, L. and Cruces, G.: Desigualdad en Argentina: una revisión de la evidencia empírica. Desarrollo Económico 192(48), 392-437 (2009)

Gasparini, L. and Sosa Escudero, W.: Assessing Aggregate Welfare: Growth and Inequality in Argentina. Cuadernos de Economía 38(113), 49-71 (2001)

Giddens, A.: The Class Structure of the Advanced Societies. Hutchinson, London (1981)

Gigliarano, C. and Muliere, P.: Measuring Income Polairzation Using an enlarged Middle Class, Working Paper 271, ECINEQ (2012)

Goldthorpe, J., McKnight A.: The Economic Basis of Social Class. CASE Papers No. 80, Centre for Analysis of Social Exclusion, London School of Economics (2004)

León, A.: Las Clases Medias en América Latina. Comisión Económica para América Latina y el Caribe, mimeo (2009)

Levy, F.: The Middle Class: Is it Really Vanishing? The Brooking Review (Fall), 17-21 (1987)

Loayza N., Rigolini J., Llorente G.: Do Middle Classes Bring Institutional Reforms?. Policy Research Working Paper 6015, The World Bank (2012)

López Calva L.F., Ortiz-Juárez, E.: A Vulnerability Approach to the Definition of the Middle Class. Policy Research Working Paper 5902, The World Bank (2011)

Ravallion, M.: The Developing World's Bulging (but Vulnerable) 'Middle Class'. World Bank Research Working Paper 4816 (2009)

Sen, A.: Poverty: an Ordinal Approach to Measurement, Econometrica, 52(3), 761-6 (1976)

Sosa Escudero W. and Petralia, S.: 'I Can Hear the Grass grow': The Anatomy of Distributive Changes in Argentina. CEDLAS Working Paper 106 (2010)

Thurow, L. C.: A Surge in Inequality. Scientific American 256, 30-37 (1987)

Tornarolli, L.: The Evolution of the Middle Class in Latin America. International Policy Centre for Inclusive Growth Working Paper 128 (2014)

Zhang, X. and Kanbur, R.: What Difference Do Polarization Measures Make? An Application to China. The Journal of Development Studies 37(3), 85-98 (2001)

Appendix

Table A.1. Size of the Poor, the Middle and the Upper Class, Argentina 1991-2012

	CENTRAL TENDENCY (CT)			POLARIZATION (PLZ)			RAVALLION (RA)			BANERJEE & DUFLO (BD)			FLAGSHIP			BARRO & EASTERLY			SOSA & PETRALIA			BIRDSALL (BIR)		
	(CT)			(PLZ)			(RA)			(BD)			(FL)			(BE)			(SP)			(BIR)		
Year	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper
1991	34.1	27.3	38.6	49.7	35.1	15.3	5.4	65.4	29.2	5.4	53.1	41.4	58.6	39.3	2.2	19.9	60.2	20.0	31.5	58.6	10.0	58.6	36.9	4.5
1992	38.8	24.7	36.5	49.8	35.1	15.2	4.6	60.5	34.9	4.6	48.0	47.4	52.6	44.9	2.5	18.0	62.0	20.0	24.9	65.1	10.0	52.6	42.4	5.0
1993	35.4	26.2	38.4	47.2	32.8	20.1	6.3	57.4	36.3	6.3	43.9	49.8	50.2	47.2	2.6	20.0	60.1	20.0	24.5	65.6	10.0	50.2	44.9	4.9
1994	35.9	25.1	39.0	49.7	35.9	14.3	5.7	59.8	34.5	5.7	46.9	47.4	52.6	44.6	2.8	19.9	60.1	20.0	26.6	63.6	9.8	52.6	42.4	5.0
1995	37.9	23.9	38.1	51.1	34.2	14.7	8.3	61.8	30.0	8.3	48.8	43.0	57.0	39.9	3.1	19.3	60.7	20.0	33.8	56.6	9.7	57.0	38.0	5.0
1996	38.2	22.0	39.5	51.9	33.1	15.1	9.7	59.5	30.9	9.7	48.1	42.2	57.8	39.7	2.5	20.0	60.1	20.0	35.9	54.1	10.0	57.8	37.2	5.0
1997	36.2	23.3	40.5	49.8	35.0	15.2	8.4	58.6	33.1	8.4	46.2	45.4	54.6	42.1	3.3	19.8	60.2	20.0	31.9	58.4	9.7	54.6	40.4	5.0
1998	36.7	21.8	41.5	49.5	35.3	15.2	9.0	56.8	34.2	9.0	45.6	45.4	54.6	41.8	3.6	20.0	60.0	20.0	33.7	56.3	10.0	54.6	40.6	4.8
1999	37.1	21.5	41.1	51.1	33.8	15.1	9.5	57.9	32.6	9.5	46.7	43.8	56.2	40.5	3.3	19.9	60.7	19.5	35.2	54.9	9.9	56.2	38.9	4.9
2000	39.3	21.0	39.4	49.7	35.1	15.2	11.1	55.3	33.6	11.1	46.8	42.1	57.9	38.9	3.2	20.0	60.1	19.9	36.5	53.5	10.0	57.9	37.2	5.0
2001	39.2	19.2	41.2	49.7	36.4	13.9	16.0	54.3	29.6	16.0	45.8	38.1	61.9	34.9	3.2	20.0	60.9	19.4	42.7	47.3	9.9	61.9	33.2	5.0
2002	38.5	23.2	38.3	49.4	35.5	15.1	25.7	56.8	17.5	25.7	49.5	24.7	75.3	23.3	1.5	19.9	60.3	19.8	62.7	27.7	9.6	75.3	19.9	4.8
2003	39.2	19.8	41.0	49.7	35.6	14.7	25.0	56.1	18.8	25.0	48.5	26.4	73.6	24.9	1.5	19.4	60.8	20.1	59.0	31.3	9.7	73.6	21.4	5.0
2004	37.5	22.3	40.2	49.6	35.1	15.2	13.4	58.8	27.9	13.4	48.5	38.1	61.9	36.2	1.9	19.8	60.2	20.0	41.1	48.9	10.0	61.9	33.1	5.0
2005	37.8	22.9	39.3	49.8	35.0	15.2	10.4	58.2	31.4	10.4	46.6	43.0	57.0	40.5	2.5	20.0	60.1	20.0	36.0	54.1	9.9	57.0	38.0	5.0
2006	38.5	21.9	39.5	49.7	35.0	15.2	8.1	53.7	38.2	8.1	41.7	50.2	49.8	46.8	3.4	20.0	60.0	20.0	30.4	59.6	10.0	49.8	45.2	5.0
2007	37.6	22.7	39.7	44.7	40.3	14.9	8.0	54.8	37.2	8.0	42.8	49.2	50.8	46.4	2.8	19.1	61.0	20.0	26.6	63.4	10.0	50.8	44.2	5.0
2008	37.5	23.3	39.2	44.7	35.0	20.2	6.9	53.2	39.9	6.9	40.5	52.6	47.4	49.8	2.7	20.0	60.2	19.9	16.7	73.5	9.9	47.4	47.6	5.0
2009	35.7	25.2	39.1	44.8	35.0	20.2	6.1	50.8	43.1	6.1	37.9	56.1	43.9	52.6	3.5	20.0	60.1	20.0	13.9	76.1	10.0	43.9	51.1	5.0
2010	35.3	25.9	38.8	44.0	35.8	20.2	4.6	50.3	45.1	4.6	36.6	58.8	41.2	55.3	3.5	19.9	60.1	20.0	10.1	80.0	10.0	41.2	53.8	5.0
2011	35.6	26.6	37.8	44.6	35.2	20.1	3.1	47.3	49.6	3.1	34.2	62.7	37.3	58.8	3.9	20.0	60.1	20.0	6.8	83.2	10.0	37.3	57.7	5.0
2012	36.1	25.1	38.8	44.7	35.1	20.2	4.6	47.0	48.5	4.6	33.9	61.5	38.5	58.2	3.3	19.9	60.1	20.0	8.8	81.3	9.9	38.5	56.5	5.0

Source: own estimations based on EPH
Notes: each cell represents the percentage of individuals classified into a particular group (poor, middle, upper class) by the corresponding measure for each year.

Table A.2. Mean Income of the Poor, the Middle and the Upper Class, Argentina 1991-2012

	CENTRAL TENDENCY (CT)			POLARIZATION (PLZ)			RAVALLION (RA)			BANERJEE & DUFLO (BD)			FLAGSHIP			BARRO & EASTERLY			SOSA & PETRALIA			BIRDSALL (BIR)		
	(CT)			(PLZ)			(RA)			(BD)			(FL)			(BE)			(SP)			(BIR)		
Year	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper
1991	103.7	209.7	598.8	130.0	320.0	962.9	43.9	180.7	696.5	43.9	155.8	575.5	145.4	476.5	2358.7	78.4	231.8	842.7	99.3	299.3	1173.3	145.4	440.6	1678.6
1992	122.8	254.6	666.9	144.4	368.3	1008.4	35.4	188.7	682.7	35.4	161.2	580.2	150.3	501.8	1973.5	80.8	259.9	890.5	95.7	323.5	1198.9	150.3	465.5	1554.9
1993	113.9	246.9	643.9	138.5	341.0	873.6	37.0	190.8	663.0	37.0	160.5	561.7	144.9	487.4	1912.0	79.0	269.4	874.2	89.9	326.3	1168.1	144.9	455.3	1539.7
1994	109.9	237.9	636.6	138.0	366.0	1027.8	37.1	186.0	678.5	37.1	157.0	573.1	144.0	484.3	1990.1	76.9	258.5	884.2	90.8	321.5	1208.0	144.0	453.0	1593.5
1995	98.9	220.8	645.9	123.8	343.2	1059.8	32.3	182.3	738.5	32.3	153.5	603.2	135.9	494.6	2001.2	62.6	233.7	912.6	91.2	322.6	1276.2	135.9	462.8	1674.9
1996	95.1	215.1	625.1	119.7	346.6	1017.7	32.9	178.5	715.1	32.9	151.6	601.3	131.7	504.9	2148.9	58.4	233.6	891.2	89.2	325.8	1217.3	131.7	464.8	1618.7
1997	96.5	222.3	648.7	124.4	361.6	1055.7	33.8	182.3	724.3	33.8	152.8	607.4	134.6	506.6	1902.6	62.6	249.7	932.0	87.4	338.0	1274.7	134.6	480.1	1641.7
1998	95.2	216.9	677.2	120.9	367.7	1141.1	36.1	178.2	757.9	36.1	150.7	642.2	131.7	516.6	2118.9	61.8	251.4	1001.3	89.3	347.4	1379.5	131.7	497.0	1879.9
1999	96.0	216.4	640.9	121.9	358.7	1071.9	36.9	178.9	730.6	36.9	152.0	617.9	132.5	505.8	2002.9	61.4	244.6	937.4	90.6	338.6	1290.3	132.5	478.5	1715.0
2000	92.1	217.8	656.2	111.2	349.7	1066.4	36.3	171.2	718.8	36.3	150.2	630.9	128.3	519.3	1975.1	55.1	236.0	936.4	85.5	338.8	1278.2	128.3	489.9	1685.0
2001	73.6	182.1	607.3	91.3	326.0	1081.1	32.2	167.8	735.8	32.2	145.4	636.3	116.0	518.6	1905.6	39.8	212.0	899.3	78.6	329.5	1256.4	116.0	484.6	1646.1
2002	50.2	130.8	439.9	63.8	212.2	737.0	34.7	157.7	683.6	34.7	138.3	568.4	102.9	482.4	1904.6	28.3	142.2	636.3	81.8	284.7	916.2	102.9	409.7	1227.3
2003	53.0	129.9	440.5	65.8	226.7	773.1	37.7	158.3	683.7	37.7	137.9	570.0	103.8	487.0	1952.5	32.2	146.2	642.1	78.8	286.3	936.5	103.8	414.0	1241.6
2004	78.3	189.5	552.8	99.9	306.4	895.5	35.5	172.6	672.6	35.5	148.3	569.6	124.0	489.9	2059.2	48.1	209.6	791.1	84.6	310.5	1073.6	124.0	441.1	1422.6
2005	93.4	216.7	615.9	116.8	341.3	979.9	37.9	180.2	694.7	37.9	152.8	585.3	131.8	496.1	2015.0	58.9	236.3	866.0	90.0	328.2	1180.7	131.8	457.4	1557.2
2006	109.0	253.2	691.6	134.3	393.9	1089.9	35.2	184.2	704.7	35.2	153.7	605.4	134.5	510.4	1933.5	68.5	273.1	966.6	91.8	357.8	1297.0	134.5	487.2	1684.3
2007	104.4	246.3	677.5	120.5	366.9	1085.3	36.2	183.1	701.4	36.2	153.1	601.0	134.7	504.2	2197.4	65.3	265.3	949.6	81.0	336.7	1292.6	134.7	473.8	1731.1
2008	116.4	263.9	692.4	133.4	359.4	943.9	37.6	188.3	686.0	37.6	157.1	590.2	139.8	509.5	2056.9	75.3	287.0	948.3	66.8	327.5	1267.4	139.8	480.3	1643.4
2009	123.4	280.6	721.5	146.1	383.5	972.7	36.5	194.0	685.9	36.5	161.4	594.5	144.1	512.8	1823.6	81.3	308.0	975.7	63.0	341.5	1279.8	144.1	493.4	1630.8
2010	134.6	292.8	734.6	156.7	393.0	985.1	40.5	197.4	681.3	40.5	163.3	590.1	149.7	512.4	1824.8	93.9	319.7	989.0	63.1	342.1	1290.7	149.7	493.7	1627.4
2011	148.7	320.6	789.8	173.2	422.8	1045.7	38.2	199.0	688.8	38.2	163.9	605.6	153.5	520.8	1887.1	105.6	343.6	1047.1	60.5	355.8	1367.4	153.5	507.8	1737.6
2012	140.2	315.4	749.0	164.3	419.4	989.0	35.2	197.7	672.1	35.2	161.6	591.4	146.7	520.1	1833.5	94.1	338.1	992.4	56.5	353.3	1292.2	146.7	500.6	1619.2

Source: own estimations based on EPH

Notes: each cell represents the mean income of individuals classified into a particular group (poor, middle, upper class) by the corresponding measure for each year.

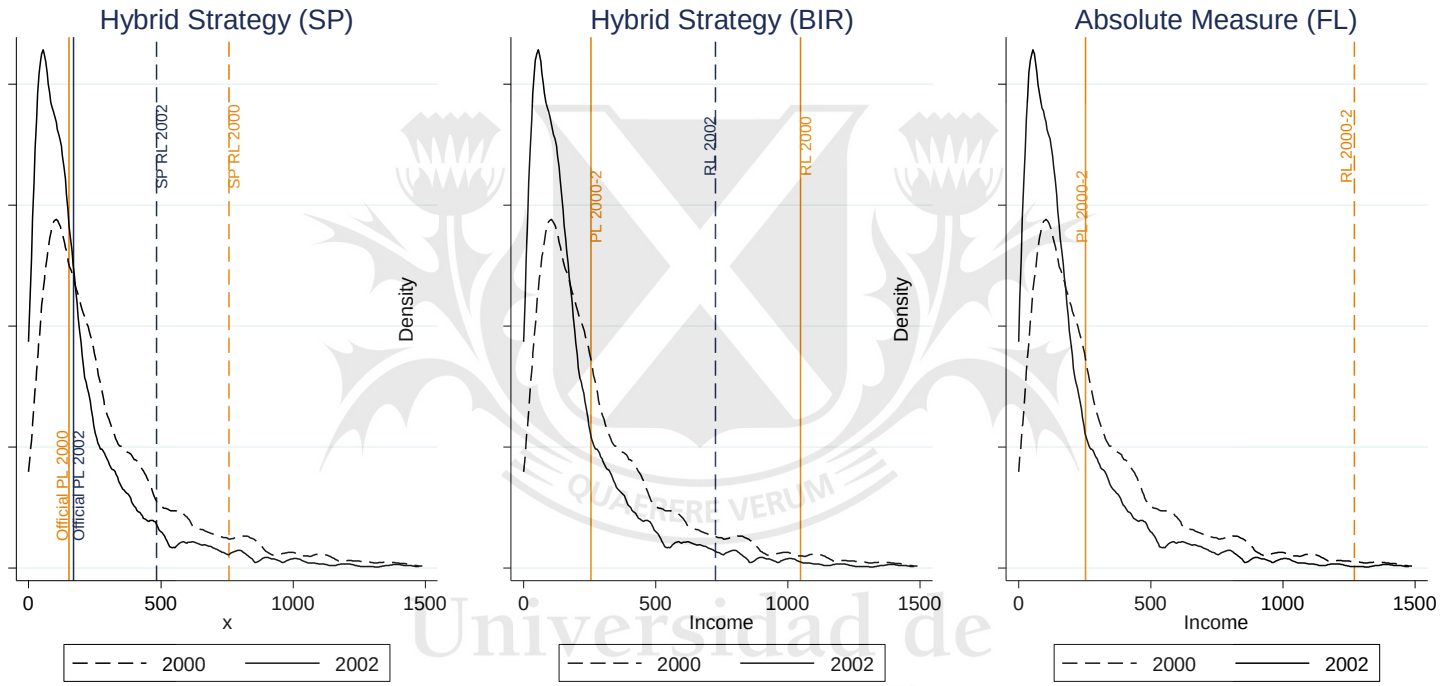
Table A.3. Mean Income Share of the Poor, the Middle and the Upper Class, Argentina 1991-2012

	CENTRAL TENDENCY (CT)			POLARIZATION (PLZ)			RAVALLION (RA)			BANERJEE & DUFLO (BD)			FLAGSHIP (FL)			BARRO & EASTERLY (BE)			SOSA & PETRALIA (SP)			BIRDSALL (BIR)		
Year	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper	Poor	Middle	Upper
1991	10.9	17.7	71.4	19.9	34.7	45.4	0.7	36.5	62.7	0.7	25.6	73.7	26.3	57.8	15.9	4.8	43.0	52.3	9.6	54.1	36.2	26.3	50.3	23.4
1992	13.5	17.8	68.8	20.3	36.5	43.2	0.5	32.3	67.3	0.5	21.9	77.7	22.3	63.6	14.1	4.1	44.7	50.5	6.7	59.5	33.8	22.3	55.7	21.9
1993	11.4	18.3	70.2	18.5	31.7	49.8	0.7	31.1	68.3	0.7	20.0	79.3	20.7	65.3	14.1	4.5	45.9	49.7	6.2	60.7	33.0	20.7	58.0	21.3
1994	11.3	17.2	71.5	19.8	37.8	42.4	0.6	32.0	67.4	0.6	21.2	78.2	21.8	62.2	16.0	4.4	44.7	50.9	6.9	58.8	34.2	21.8	55.3	22.9
1995	11.1	15.7	73.2	18.8	34.9	46.3	0.8	33.4	65.8	0.8	22.2	77.0	23.0	58.6	18.4	3.6	41.3	54.3	9.1	54.2	36.6	23.0	52.2	24.8
1996	10.6	15.0	74.8	18.8	34.7	46.5	1.0	32.2	66.9	1.0	22.1	76.9	23.1	60.8	16.1	3.5	42.5	54.0	9.7	53.5	36.8	23.1	52.4	24.5
1997	10.0	14.9	75.2	17.7	36.2	46.1	0.8	30.6	68.6	0.8	20.2	79.0	21.0	61.1	17.9	3.5	42.4	53.5	8.0	56.5	35.5	21.0	55.6	23.4
1998	9.6	13.0	77.4	16.5	35.7	47.9	0.9	27.9	71.2	0.9	18.9	80.2	19.8	59.5	20.7	3.4	41.5	55.1	8.3	53.9	37.9	19.8	55.6	24.7
1999	9.8	14.3	76.4	18.0	35.2	46.8	1.0	30.0	69.0	1.0	20.5	78.4	21.6	59.4	19.0	3.5	43.0	54.4	9.2	53.9	36.8	21.6	53.9	24.5
2000	10.2	14.1	76.1	16.3	36.0	47.7	1.2	27.8	71.0	1.2	20.6	78.2	21.8	59.4	18.8	3.2	41.7	55.3	9.2	53.3	37.6	21.8	53.5	24.6
2001	8.9	11.1	79.6	14.4	37.7	47.9	1.6	29.0	69.4	1.6	21.2	77.2	22.8	57.6	19.6	2.5	41.0	58.1	10.7	49.6	39.7	22.8	51.1	26.1
2002	8.9	13.9	77.2	14.5	34.6	51.0	4.1	41.0	54.9	4.1	31.4	64.5	35.5	51.4	13.1	2.6	39.3	58.6	23.5	36.2	40.3	35.5	37.5	27.0
2003	9.2	11.3	79.5	14.4	35.5	50.1	4.2	39.2	56.7	4.2	29.5	66.3	33.7	53.5	12.9	2.8	39.0	60.0	20.5	39.4	40.1	33.7	39.1	27.2
2004	10.0	14.4	75.6	16.9	36.6	46.5	1.6	34.5	63.9	1.6	24.5	73.9	26.1	60.3	13.6	3.2	42.9	53.9	11.8	51.7	36.5	26.1	49.7	24.2
2005	10.8	15.2	74.0	17.8	36.5	45.7	1.2	32.1	66.7	1.2	21.8	77.0	23.0	61.4	15.6	3.6	43.4	53.0	9.9	54.3	35.8	23.0	53.2	23.8
2006	11.3	15.0	73.7	18.0	37.2	44.8	0.8	26.7	72.5	0.8	17.3	81.9	18.1	64.5	17.5	3.7	44.2	52.1	7.5	57.5	35.0	18.1	59.4	22.5
2007	10.8	15.4	73.8	14.8	40.6	44.6	0.8	27.6	71.6	0.8	18.0	81.2	18.8	64.2	17.0	3.4	44.1	52.2	5.9	58.6	35.5	18.8	57.6	23.7
2008	11.6	16.3	72.1	15.8	33.4	50.7	0.7	26.6	72.7	0.7	16.9	82.4	17.6	67.4	15.0	4.0	45.9	50.4	3.0	63.9	33.2	17.6	60.7	21.7
2009	11.1	17.9	71.1	16.5	33.9	49.7	0.6	24.8	74.6	0.6	15.4	84.0	16.0	68.0	16.1	4.1	46.6	49.4	2.2	65.6	32.2	16.0	63.5	20.5
2010	11.6	18.6	69.8	16.9	34.4	48.7	0.5	24.3	75.3	0.5	14.6	84.9	15.1	69.4	15.5	4.6	47.0	48.4	1.6	67.0	31.5	15.1	65.0	19.9
2011	12.1	19.6	68.3	17.7	34.1	48.2	0.3	21.6	78.2	0.3	12.8	86.9	13.1	70.1	16.8	4.8	47.3	48.1	0.9	67.8	31.3	13.1	67.1	19.8
2012	12.0	18.8	69.1	17.5	35.0	47.5	0.4	22.1	77.5	0.4	13.0	86.6	13.4	72.0	14.6	4.5	48.3	47.3	1.2	68.3	30.5	13.4	67.3	19.2

Source: own estimations based on EPH

Notes: each cell represents the share of mean income held on average by individuals classified into a particular group (poor, middle, upper class) by the corresponding measure for each year.

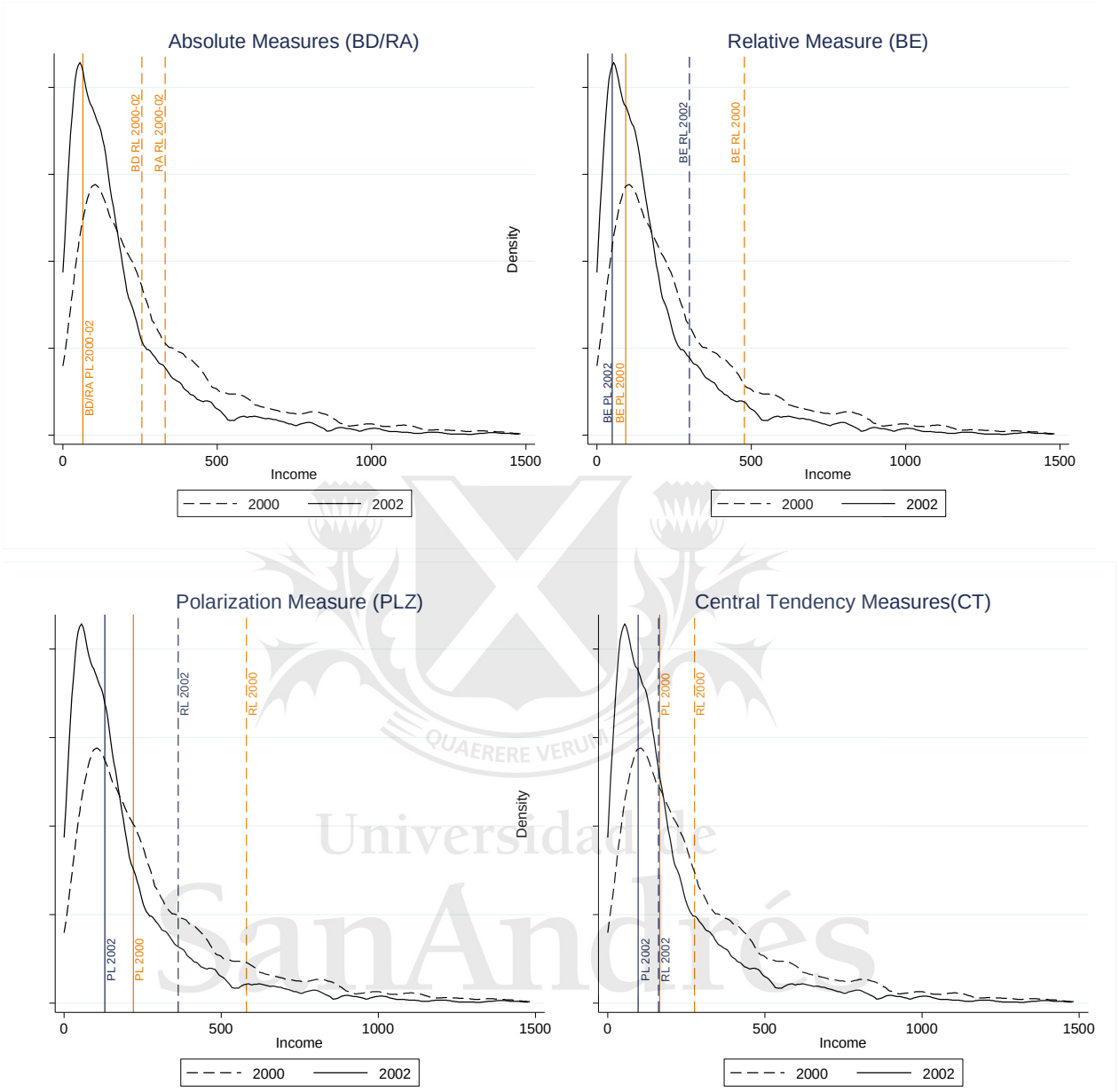
Figure A.1. Lower and upper thresholds during 2001-02 crisis (SP/BIR/FL)



Source: own calculations based on EPH

Notes: each figure shows the kernel density of income for 2001 (solid line) and 2002 (dashed line). Poverty and Richness lines for each measure are included in each figure.

Figure A.2. Lower and upper thresholds during 2001-02 crisis (BD/RA/BE/PLZ/CT)



Source: own calculations based on EPH

Notes: each figure shows the kernel density of income for 2001 (solid line) and 2002 (dashed line). Poverty and Richness lines for each measure are included in each figure.

CHAPTER II. Measuring the middle class in many dimensions: Argentina 2004-2014*

Abstract

Chapter I showed middle class studies have gained relevance in the economic literature yet were far from conceptual and methodological agreements regarding its measurement. A unidimensional identification of the Argentinean middle class was carried out to assess the implications of different definitions based solely on income. Two limitations of this unidimensional identification were highlighted: the difficulties in setting an upper bound as well as disavowing the rich literature coming from the sociological and political theory realms that point to other dimensions as key in defining the middle class. In order to address those difficulties, in this Chapter we present a new multidimensional approach for identifying the middle class based on multivariate quantiles. The procedure is employed to multidimensionally identify the Argentinean middle class and track its evolution across the 2004-2014 period, characterizing its performance and main features. Furthermore, a variable selection exercise is also proposed in order to identify the most relevant variables in terms of welfare and allocation to groups (poor, middle and upper class).

JEL classification: D3, I3, D6
Keywords: middle class, distribution, Argentina

* This article is joint work with Walter Sosa Escudero (UdeSA, CEDLAS-Universidad de La Plata and CONICET) and Marcela Svarc (UdeSA and CONICET).

1. Introduction

As stated in Chapter I, the economic literature has recently devoted increasing importance to the middle class (Atkinson and Brandolini, 2011; Ravallion, 2009; Birdsall *et al.*, 2000). However, it is still lagging behind as compared to poverty or inequality analysis, especially for the case of Argentina. Given this relative void, Chapter I contributed to the literature by making a characterization of the Argentinean middle class during the last two decades based on a unidimensional identification. The economic performance of this group was tracked using a number of different measures based on income.

This very simple exercised helped to clarify some crucial aspects of measuring a group such as the middle class but also showed the difficulties found in defining it by only focusing on income. In particular, we highlighted three limitations of this unidimensional identification of the middle class. In the first place, the poverty literature has already established the idea that only multidimensional measures may actually capture well-being. This becomes even more relevant when focusing on the middle class. On the one hand, even though certain agreement may be reached on establishing a lower bound in terms of income, it is far less obvious how to settle an upper threshold. On the other hand, this implies disavowing the large and rich literature coming from the sociological and political theory realms that point to other dimensions as key in defining the middle class: the occupational structure, the level of education, wealth, etc[15].

In light of these difficulties, and mirroring the path followed by poverty and inequality measurement, this chapter proposes a new multidimensional approach to identify the middle class. Indeed, it is interesting to reproduce the exercise of tracking the evolution of the Argentinean middle class pursued in Chapter I but departing from a multidimensional identification of this group, rather than a unidimensional one. In particular, we would like to assess its economic performance across the last decade as well as characterize its main features.

We propose to extend a unidimensional relative measure of the middle class based on quantiles of the income distribution to a multidimensional setting. Therefore, we need to resort to a definition of multivariate quantiles. Even though the literature offers several approaches, none of them are suitable for our particular problem: their equivariance does not internalize the nature of welfare variables that are increasing in order and they do not ensure that the resulting region will contain the desired mass of the population.

[15] This goes as further back as Max Weber (1905) and was more recently incorporated to the economic literature (see Atkinson and Brandolini, 2011)

We thus propose a new approach to define multivariate quantiles that is suitable for identifying the middle class. The procedure is as follows. Firstly, a new multidimensional welfare measure is derived through the construction of multivariate quantiles. Departing from multivariate data, a unidimensional well-being index is defined by projecting the data of the original space on a growth direction of increasing welfare which is based on principal components analysis. In particular, the growth direction is derived from the first principal component. Secondly, quantiles are defined over this index, establishing a lower and an upper threshold. These quantiles are shown to truly identify the middle class multidimensionally. Furthermore, we present a new approach to reduce the dimensionality of welfare. Based on the blinding strategy proposed by Fraimant *et al*. (2008) and on the selection variable procedure in Edo *et al.* (2016), we select the smallest subset of variables belonging to the original space that are able to reproduce both the unidimensional well-being index and the partition into three groups induced by the multidimensional quantiles. In this way, we are able to identify how many and which are the dimensions relevant to define welfare as well as to distinguish the middle class from the poor and the upper class.

The approach presented overcomes previous methods' shortcomings. To start off, some authors claim to study the middle class in multidimensional terms but fail to provide a truly multidimensional identification. For instance, Davis and Huston (1992) and Gayo (2013) use income thresholds to identify the middle class and then analyze this group's evolution by expanding the analysis to other dimensions. In second place, previous approaches that actually multidimensionally identify the middle class do not ensure that the group selected actually belongs to the central region of the multivariate distribution (Gigliarano and Mossler, 2009). Thirdly, other attempts to multidimensionally identify the middle class are not able to provide an exhaustive classification of the individuals not belonging to this group. For instance, Gigliarano and Mossler (2009) identify the middle class multidimensionally yet do not provide a sound classification of the poor and the upper class. Finally, none of the mentioned approaches proposes a method to reduce multidimensionality and identify how many and which are the key variables needed to identify the middle class.

This new approach is used to multidimensionally identify the Argentinean middle class during the 2004-2014 period, so as to extend the analysis made in Chapter I to a multivariate setting. We depart from 19 variables associated to different aspects of well-being: per capita family income; sources of income, property and wealth; employment and education; dwelling characteristics and having a domestic employee. Once the Argentinean middle class is defined in multidimensional terms, we are able to track its economic performance across that period as well as characterize its

main features. Furthermore, we explore how many and which are the variables that may reproduce the results obtained with the original space.

The rest of the article is organized in the following way: Section 2 describes the theoretical and empirical approach based on the α-quantile region definition orientated by a growth direction. Section 3 applies the new approach to identify the Argentinean middle-class under the 2004-2014, tracks its economic evolution and describes its main features as well as shows which are the key variables that define welfare and grouping allocation. Section 4 concludes.

2. Middle Class and Multidimensional Identification

This section aims at developing an approach that is able to identify the middle class from a given set of multidimensional well-being information. We rely on a relative measure of the middle class. Indeed, we aim at extending to a multivariate setting unidimensional relative measures based on quantiles of the income distribution[16].

This implies extending the univariate concept of α-quantile to the the multivariate setting. Such as in the univariate concept, the middle class will be defined as the subset of individuals within a lower bound that separates the poor from the middle class, and an upper bound that separates it from the upper class. These thresholds, however, will be defined in terms of a multivariate notion of quantiles.

The nature of the problem at hand, imposes two properties on the definition of the multivariate quantiles. On one hand, the multivariate α-region, $C(\alpha)$ must have mass greater than or equal to α, i.e. $P\big(\mathrm{X} \in \mathrm{C}(\alpha)\big) \geq \alpha$, since we define the middle class as a given proportion of central population. On the other hand, since our variables measure well-being, each of them has a natural increasing order. This order must be preserved by the definition stated, implying that the quantile function defined will not be equivariant. These properties are not fulfilled by previous definitions developed in the literature. In fact, even though the concept of a multivariate quantile has been largely studied[17], none of these definitions are suitable for our analysis. There are two main drawbacks. First of all, quantile functions on R^p p are desirably equivariant, that is the new quantile representation of a point x after affine transformation should agree with the original representation similarly transformed. Secondly, there are many definitions of multivariate α-quantile, most of them

[16] For instance, Barro and Easterly (2001) propose classifying as middle class those who lie between the 3rd and 8th decile of the income distribution.

[17] See for instance, Chauduri (1996), Serfling (2010), Hallin et al. (2010), Fraiman and Pateiro-Lopez (2012) and Kong and Mizera (2012).

define α-quantiles orientated by a given direction, hence considering all the unitary directions a region in the space is determined, however there is no relation between the probability of these regions and the directional α-quantiles. We therefore propose a new approach for defining multivariate quantiles that satisfies the goal of identifying the middle class. This is the subject of the following subsection.

2.1 Theoretical Approach and Empirical Model

Let X be a $p-$dimensional random vector with distribution P_x, representing aspects of social and economic well-being. The goal is to extend the univariate concept of $\alpha - quantile$ to the the multivariate setting. As mentioned in the Introduction a first goal is to determine the $\alpha - upper$ (and an $\alpha - lower$) region of the distribution. A basic monotonicity assumption is that each random variable in the multidmensional welfare space follow a a natural increasing order, that is higher levels of each of them correspond to increasing levels of well-being.

The proposal is to project the data into the direction of g_D, which denotes the *growth direction*. To attain uniqueness this direction must have unitary norm. Given the nature of the problem under anaylisis, i.e.: well-being, $g_D$should be positive coordinate wise. If there is no previous knowledge of the distribution, a natural growth direction could be $g_D = (1, \dots, 1)/\sqrt{p}$, that is, the mean of the welfare dimensions. In our setting (multidimensional welfare), different variables may have different weights. In order to maximize this variance, we propose to define g_D as the first principal component.

Let $B = \{X \in \mathrm{R^p}: \|\mathrm{X}\| = 1\}$, $g_D \in B$, if all the coordinates are positive. We denote $Y_D = \langle X, g_D \rangle$, the projection of X with respect to g_D. Following Fraiman and Pateiro-Lopez (2012), let

$$\tilde{Q}(\alpha, g_D) = inf_{t \in \mathrm{R}} \{F_{\langle X - E(X), g_D \rangle}(t) \geq \alpha\} \quad (1)$$

Where

$$F_{\langle X - E(X), g_D \rangle}(t) = P(\langle X - E(X), g_D \rangle \leq t) \quad (2)$$

Then the $\alpha - quantile$ in the direction of g_D is given by

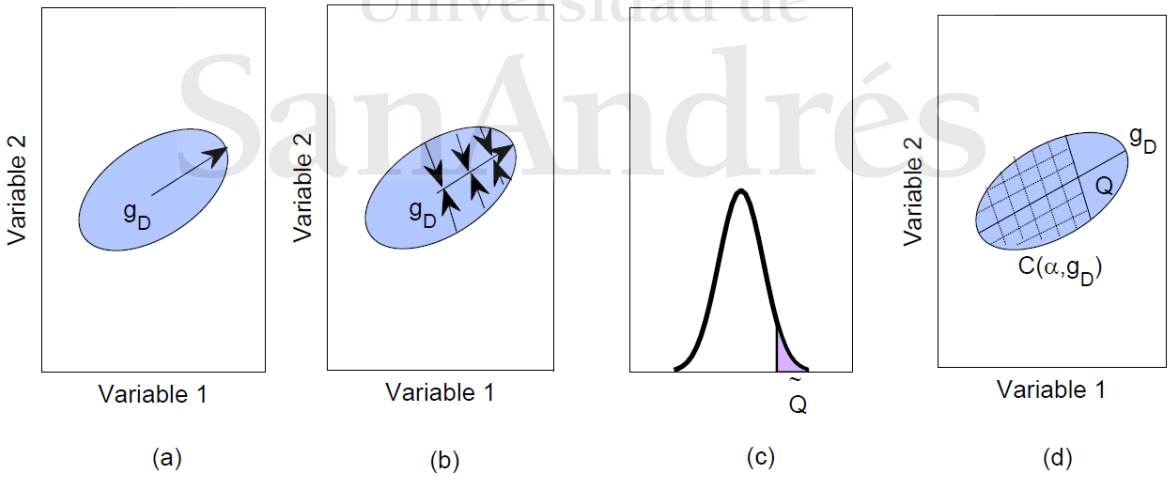
$$Q(\alpha, g_D) = \tilde{Q}(\alpha, g_D) g_D + E(X) \quad (3)$$

Then we define the $\alpha - quantile$ region as

$$C(\alpha, g_D) = \{x \in \mathrm{R^p}: \langle x - \mathrm{E(X)}, g_D \rangle \leq \tilde{Q}(\alpha, g_D)\} \quad (4)$$

Figure 1 presents a sketch of the procedure proposed. The intuition is as follows. We first project the data in a growth direction (g_D) which is determined by the first principal component (see Figure 1, panels (a) and (b)). This in turn ensures that the growth direction complies with the two basic properties required: it has unitary norm and is positive coordinate wise. The latter implies that the projection will lie in the first quadrant. This projection of the data, i.e $Y_D = \langle X, g_D \rangle$, may be thought of as a uni-dimensional "well-being indicator". Equation (2) represents the cumulative distribution function associated to a standardized version of this uni-dimensional well-being indicator, which may be appreciated in panel (c) of Figure 1. This function evaluated at point *t* reflects the percentage of observations of the projection that have a value lower than or equal to *t*. Equation (1) represents the well-being indicator quantile, $\tilde{Q}(\alpha, g_D)$. That is, it denotes the value at which the cumulative distribution function is larger than (or equal to) α. Equation (4) determines the $\alpha - quantile$ region, as may be intuitively appreciated in panel (d), Figure 1. That is, the region within which all observations when projected are lower than the well-being indicator quantile $\tilde{Q}(\alpha, g_D)$. It is clear that this region is bounded by the hyperplane that is orthogonal to $g_D$and that contains the point $Q(\alpha, g_D)$[18].

Figure 1. The multivariate quantile approach



Once the theoretical approach has been described, we now turn to the empirical model describing the empirical counterpart of the multivariate quantiles defined.

[18] Edo *et al.* (2016) prove proper coverage probability to ensure that the number of observations that falls below $\tilde{Q}(\alpha, g_D)$ is the same for the unidimensional well-being indicator and for X^p.

Let $X_1, \dots, X_n$ be a random sample of vectors with distribution P_X and denote by P_n its empirical distribution. In order to define the empirical counterpart of the $\alpha - quantile$ on the direction of g_D, we first need to define the empirical expression for (1)

$$\tilde{Q}_n(\alpha, g_D) = inf_{t \in \mathrm{R}}\{F_{n,\langle X-\bar{X}, g_D\rangle}(t) \geq \alpha\}, \qquad (5)$$

where

$$F_{n,\langle X-\bar{X}, g_D\rangle}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\langle X-\bar{X}, g_D\rangle \leq t\}}. \quad (6)$$

Then the empirical expression for (3) is

$$\hat{Q}_n(\alpha, g_D) = \tilde{Q}_n(\alpha, g_D) g_D + \bar{X}. \quad (7)$$

The empirical counterpart for the $\alpha - quantile$ region is

$$C_n(\alpha, g_D) = \left\{x \in \mathrm{R}^{\mathrm{p}}: \langle x - \overline{\mathrm{X}}, g_D\rangle \leq \tilde{Q}_n(\alpha, g_D)\right\} \quad (8)$$

Therefore, the empirical approach requires to first project the multidimensional data in the growth direction (g_D), determined by the first principal component. The empirical unidimensional welfare index is thus constructed, i.e. $\widehat{Y_D}$. Then a multivariate quantile is determined ($\tilde{Q}_n(\alpha, g_D)$), that is, the value of the index that ensures that a proportion α of the observations lies above that cut-off point. Finally, the approach goes back to the original multidimensional space to define the $\alpha - quantile$ region, which includes all observations such that when projected have a value larger (or equal to) $\tilde{Q}_n(\alpha, g_D)$[19].

In order to determine a middle class in relative terms, this procedure is repeated twice. In effect, the middle class as defined in relative terms represents the group that belongs to some pre-determined quantiles of the multivariate distribution. Therefore, an upper and a lower quantile need to be determined. By construction, this will not only identify the middle class but also the poor and the upper class, implicitly defining a "poverty" and a "richness" threshold.

The approach presented overcomes previous methods' shortcomings. First of all, it provides a truly multidimensional identification of the middle class. Several authors claim to study the middle-class

[19] See Edo *et al.* (2016) for proof of strong convergence of the empirical to the theoretical multidimensional quantile regions.

multidimensionally (see for instance Davis and Huston, 1992 and Gayo, 2013) but they define the middle class in terms of the income and use several dimensions for the subsequent analysis. This truly multidimensional identification relies on the projection of the variables in the original multidimensional space. Furthermore, determining the absolute value of the first principal component as the growth direction we are also maximizing the information contained in the original multidimensional space. Secondly, it ensures that the group identified as the middle class actually belongs to the central region of the multivariate distribution given the fact that variables represent welfare and thus have a natural increasing order. This is not an obvious result. Gigliarano and Mosler (2009) propose a multidimensional identification of the middle class through the definition of an ellipsoid that covers at least a given proportion of population and has minimum volume among all such ellipsoids. Such an approach cannot guarantee the identification a subset in the central region of the distribution. It will tend to capture the most dense region, and there is no reason to assume that it will contain the central observations. In contrast, the multivariate quantile setting proposed in this paper defines by construction that the middle class will belong to the central region of the multivariate distribution when taking into account the natural ordering implied by the welfare variables. Thirdly, our proposal generates an exhaustive classification of individuals along an increasing well-being index. That is, not only the middle class is defined, but also the poor and the upper class. This allows for comparisons between our group of interest and other multidimensionally identified groups. Once again, this is not an obvious result. The approach suggested by Gigliarano and Mosler (2009) to multidimensionally identify the middle class does not ensure this. They define the middle class as a convex central region, typically a ball with center in the multidimensional mean of the attributes and a varying radius determining a region containing a given proportion (for instance, 50%) of the observations. The observations that are not classified as "middle class" are left unclassified. Indeed, even if it could be possible to argue that some of the observations are clearly worse-off that those belonging to the middle class (bounded to be "poor") and some others are clearly better-off (the "upper class"), we are still faced with a relevant region for which classification is not trivial. On the contrary, our approach guarantees that the group defined as the middle class actually lies in the middle of the multivariate distribution. This is ensured both by the generation of a unidimensional welfare index as well as the identification of the different groups (poor, middle class, upper class) through multivariate quantiles.

2.2 Variable Selection Analysis

Section 2.1 defines a multivariate quantile function for any arbitrary multivariate notion of welfare. An important question is whether all initial variables are equally important, since it might be the

case that some variables only add noise or can be appropriately captured by other variables. In particular, we are interested in reducing the number of dimensions in two different ways. On one hand, we want to know which is the minimum set of variables that may reproduce fairly accurately the unidimensional welfare index. That is, which and how many are the variables needed to construct a unidimensional welfare index as similar as possible to the one projected from all of the variables in the original space. On the other hand, it is interesting to show which are the key variables for distinguishing among the different groups: the poor, the middle and the upper class. For instance, we are interested in zooming into the particular poor/middle-class divide and identify which (and how many) are the relevant variables that define it. A 'relevant' variable is defined in terms of the amount of individuals that would be misclassified (i.e. considered poor when they were identified as middle class when using the original space). Larger misclassifications imply that the variable is more relevant. Our aim is to find the minimum set of variables that ensure that this misclassification is fairly small. This analysis is carried out for both divides: poor/middle class and middle/upper class. We therefore carry out three variable selection procedures that identify the minimum subset of variables such that: *(i)* the unidimensional welfare index is reproduced fairly accurately; *(ii)* the classification of an observation into the poor or the middle class is by and large replicated; *(iii)* the classification of an observation into the upper or the middle class is by and large replicated.

The procedure for reducing dimensionality in all three cases is based on a blinding strategy that eliminates unnecessary variables developed in Fraiman *et al*. (2008). Their approach is based on the idea of blinding unnecessary non-informative or redundant variables. We will discuss the main intuitive ideas behind the procedure, the technical counterpart may be found in Fraiman *et al.* (2008). For simplicity, suppose there are only two variables in the original space, X and Y. If our unidimensional welfare index were obtained based only on X, Y is redundant if: a) does not add information (it is captured by X) or b) it is not related to X but leaves the index unaltered when incorporated in the projection (i.e., it only adds 'noise'). If Y is redundant, then the projection will remain unaltered if Y is replaced by its best prediction based on X, that is, its conditional expectation[20]. For illustration, if correlation between X and Y were 1, then Y would be replaced by X. If it were 0, Y would be replaced by a constant. The aim is then to find the smallest subset of the original space that can reproduce both the unidimensional welfare index as well as the classification generated by the poor versus middle class divide (and middle versus upper class). For this, blinded random vectors are constructed where a subset of variables of the original space are replaced by

[20] This conditional expectation is based on the r−-nearest neighbour (r-NN) estimates.

their conditional expectation. In this way, all possible random blinded vectors are constructed, based on all possible subsets of variables of the original space. This blinding strategy is common to all three variable selection exercises. Once the blinded vectors are constructed, however the procedure followed for *(i)* reproducing fairly accurately the unidimensional welfare index, is different than that carried out for *(ii)* and *(iii)*, replicating by and large the classification of the observations into the different groups.

For the case of the unidimensional welfare index, the procedure follows the method developed in Edo *et al.* (2016). Intuitively, the main idea guiding the process is to compare the projection generated by the whole set of variables in the original welfare space to the one constructed from each blinded random vector. Actually, the cumulative distribution function of both indexes are contrasted in an hypothesis test where the null hypothesis is that both distributions are equal and the alternative is that they are not. For p-values larger than 0.2, the subset of variables are claimed to represent quite accurately the welfare index based on the original space. The aim is to choose the minimum subset that ensures that both cumulative distributions are not statistically different.

The procedure for identifying the key variables in the poor/middle-class divide is different. In this case, the Fraiman *et al.* (2008) methodology is followed. We introduce here the main intuitive ideas First of all the classification of a certain individual to one of both groups (poor or middle class) is made based on the projection constructed from the original space. In a second stage the same classification is made based on one of the random blinded vectors (i.e, the projection is constructed using a subset of the original variables). We then assess the percentage of observations that are reallocated after the blinding procedure. We establish 5% as the maximum percentage of observations that could be reallocated. Following this threshold, we compute the reallocation implied by each blinded vector and identify the minimum subset of variables that comply with this rule. The procedure is repeated to zoom into the middle/upper-class divide.

3. Middle Class in Argentina 2004-2014

3.1 Data

For this purpose we rely on micro data coming from the *Encuesta Permanente de Hogares* (EPH). The survey provides quarterly information on demographic aspects, education, employment and family income as well as characteristics of the dwelling for households across the country.

Given the aim of the present analysis, we include a large set of variables in order to multidimensionally identify the Argentinean middle class. The choice of the initial set of variables

is certainly arbitrary. Even though it is based on conceptual reasons it is also limited by the information provided by the EPH which was not designed to assess well-being.

In total, we select 19 variables that to some extent we may claim that identify different dimensions of welfare. In the first place we consider family income. Even though our objective is, precisely, to transcend this dimension, it remains one of the most relevant factors that determines household welfare.

A second set of variables are incorporated, following the lines of classical economists who related class to the sources of income, property and wealth (Atkinson and Brandolini, 2011). In this sense, we incorporate variables that identify whether the family has access to income from renting some other property, from profits of a business in which they do not actively participate as well as income from financial assets. Information on ownership of the dwelling is also incorporated. This is especially important in a country such as Argentina where access to mortgage credit is very expensive. Furthermore, data on whether the household receives subsidies is included. Finally, we also contemplate consumption strategies: we include a variable that identifies whether or not the household needs to rely on installments to acquire goods.

A third set of variables addresses the concerns of the sociological and political theory literature that associated classes to labor market stratification. In the first place, we include data on whether the head of household is employed. We then move on to identify the occupation type of the household head, from unskilled employment to professional positions. In line with this, we incorporate variables on educational level of the head of household.

We also include characteristics of the household's dwelling. In particular, we concentrate on its construction materials, its access to basic services as well as whether it is located in risky areas (flood zone, slums, etc). Finally, our analysis also incorporates one additional variable not traditionally included in middle-class studies: whether the household relies on a domestic employee to take care of household chores, a common practice among Latin American countries. This group represents the original welfare space. Table A.1 in the Appendix provides a complete list of variables with a more detailed description.

The time span under analysis is 2004-2014. For each of these years, data for all four quarters are provided. Analysis are carried out independently for each quarter, implying more than forty data subsets. On average, each quarter contains around 16,500 households, summing up to around 712,000 observations for the whole period under consideration.

3.2 Middle Class multidimensional identification in Argentina 2004-2014

As stated in the previous section, we have defined a multidimensional welfare space. In particular, we have included nineteen variables related to different aspects of well-being as suggested by the literature. In order to be able to assess the multidimensional welfare of these individuals we will proceed to apply the multidimensional quantile approach explained in the previous section. In essence, our aim is to project the information contained in our original multidimensional space by way of establishing a growth direction that ensures a consistent ordering of the individuals in terms of well-being. In other words, we establish a sort of welfare index that may allow for a consistent ranking of welfare across-individuals, departing from multidimensional data that has not an obvious order.

As exposed in the previous section, the first step is to resort to principal components analysis. We apply a principal component factorization for all the quarters under analysis[21]. Results suggest that our nineteen variables across more than forty quarters may appropriately summarized by four orthogonal factors, accounting for around 80% of total variability. Our approach defines the growth direction by which the original space is projected by the first principal component[22]. Two results of the principal components analysis give strength to this procedure. On the one hand, the first principal component accounts for 30% of variability on average across quarters, which is high relative to the magnitude of our original space. On the other hand, the weights assigned to each variable within the first principal component are on average the same across *all* quarters and years, ensuring consistent comparison across periods. This is surprising given the fact that we are repeating the analysis for over forty datasets (see Figure A.1 in the Appendix).

The final result of this procedure is a well-being indicator, that allows for a consistent ranking of individuals using information coming from several dimensions of welfare. For Argentina under the period of evaluation we established the lower $\alpha - quantile$ as $\tilde{Q}(25, g_D)$, while the upper $\alpha - quantile$ is $\tilde{Q}(90, g_D)$. This implies that all observations that once projected into the unidimensional welfare indicator lie between the 25th and 90th of the multidimensional percentile are included into the middle class. That is, the $\alpha - quantile$ region is defined as the one containing all observations that once projected belong to that interval. Such a choice of percentiles (25th – 90th) is certainly arbitrary but grounded on previous literature. Edo and Sosa Escudero (2012) find that

[21] All variables are standardized before this procedure to ensure a consistent analysis.
[22] Given our data, it was necessary to take the *module* of the first principal component the growth direction. One of the loadings obtained resulted negative and we needed to ensure that the growth direction was positive coordinate wise.

defining the 90th percentile as an upper threshold works well enough for unidimensionally identifying the middle class[23].

3.3 Middle Class performance in Argentina 2004-2014

Once the Argentinean middle class has been multidimensionally identified, we proceed to characterize its performance during the period under analysis. In first place, we focus on how it has fared in economic terms. Secondly, we describe its main socio-demographic features across the period.

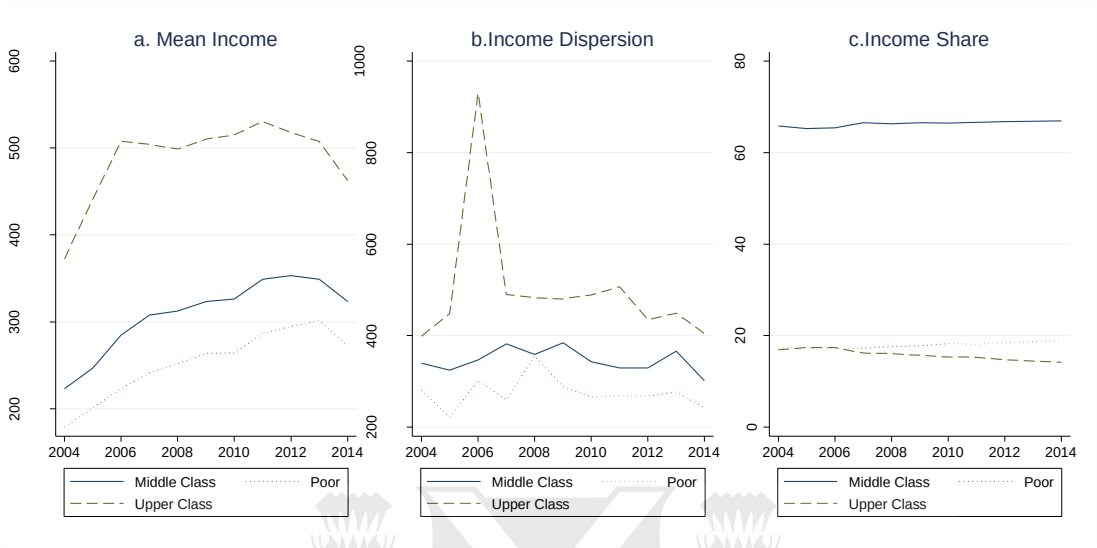
Economic Performance

The middle class measure proposed in this approach implies that the size of the group under analysis is fixed. In particular, 65% of the population by definition will be identified as the middle class following our approach. In this context, its evolution in economic terms may only be measured in terms of the path followed by some welfare indicators, such as the level of income, income share as well as its dispersion. Figure 2 shows all of the three indicators mentioned across the 2004-2014 period, for all groups identified by our method: the poor, the middle and the upper classes.

It can be observed that the middle class seems to have fared rather well across the period. In fact, the results are consistent the findings shown in Chapter I which stemmed from a unidimensional identification of the middle class. On the one hand, mean income seems to have risen steadily from 2004, showing a slight decrease in 2013 (see Figure 2a)[24]. It is worth noting, however that this is the case for both the poor and the upper class, which suggests that this indicator may be reflecting the path of the economy in general rather than the particular evolution of the middle class as a specific group. In terms of income share, it seems to have remain stable across the period. Holding around 60% of the income share this seems a positive result for the middle class, given the fact that, by definition, the group holds 65% of the population. In terms of dispersion, we also detect signs of stability. This could also be interpreted in some sense in terms of internal cohesion, which should be viewed as positive in light of maintaining levels of polarization low and stable[25].

[23] Several robustness checks were run, setting different upper and lower bounds. The conclusions presented in the following sections are not sensible to those alternative scenarios.
[24] Income has been deflated to make the different periods comparable.
[25] These results are in line with results found in Chapter 1: as from 2004 the middle class seems to have performed well in terms of income and income share.

Figure 2. Middle Class in Argentina 2004-2014, Economic Performance



Source: own calculations based on EPH.

Notes: 'Mean income' refers to the average income enjoyed by each group during each year. 'Income dispersion' reflects the average standard deviation in income for each group in each year. 'Income share' indicates the portion of total income hold by each group on average each year.

Socio-demographic characteristics

Beyond the economic evolution traced previously, it is also interesting to characterize the Argentinean middle class in terms of its demographic characteristics, how these differ from other groups and how these features may have changed over time. Table 1 shows some of these characteristics for 2004 and 2014.

On average, the middle class in Argentina shows the largest household size as compared to the other two groups, even though it is not especially large (4.4 individuals per household across the period). As expected, they show a greater share of children (around 70% of middle class households have children under 18). The head of household is generally a male, and in this dimension it clearly differs from the poor for whom the head is female for around 50% of households.

Table 1. Middle Class in Argentina 2004-2014, Socio-demographic Features

Category	Variable	Poor		Middle Class		Upper Class	
		2004	2014	2004	2014	2004	2014
Demographics	Household Size	4.1	4.1	4.5	4.4	4.3	4.2
	Number of children <18	1.3	1.3	1.8	1.7	1.5	1.4
	% of HH with children <18	51.7	54.6	71.5	70.3	68.9	67.9
	% of Female head of HH	46.1	55.1	44.6	32.4	19.8	27.2
Education	% of Head of HH with completed secondary or higher	27.8	32.7	42.9	48.9	72.8	78.3
	% of Head of HH with completed university degree	5.8	8.2	13.4	16.2	39.3	44.2
Employment	% of head of HH employed	7.7	9.0	86.0	83.6	100.0	100.0
	% of HH members employed	17.8	19.9	50.2	51.8	57.7	61.3
	% of female workers in HH	36.0	34.1	54.5	56.7	61.0	64.8
	% of head of HH in formal sector	39.7	59.8	58.1	71.2	76.6	86.3
	% of head of HH self-employed	31.9	20.5	24.3	23.3	22.2	21.1
	% of head of HH business owner	0.9	1.3	4.8	4.1	11.3	8.2
	% of head of HH in unskilled employments	49.2	41.9	19.9	16.8	3.0	4.8
	% of head of HH in high-skilled employments	1.2	5.2	23.0	24.0	47.5	52.5
Subsidies	% of HH receiving subsidies	21.9	25.5	16.6	22.5	5.2	8.0
Ownership	% of HH owners	62.8	60.4	66.2	63.8	90.8	83.2
Consumption Strategy	% of HH buying in installments	77.7	63.3	76.1	56.1	41.4	8.7
Dwelling Charachteristics	% with solid floor	73.6	77.6	73.2	81.0	88.8	90.9
	% with adequate sewage	82.4	87.3	84.2	89.2	92.9	94.1
Other	% of HH with domestic employee	3.8	3.2	5.9	4.0	19.9	11.8

Source: own calculations based on EPH.

Notes: HH stands for 'household'. '% of HH members employed' is the ratio of employed individuals within the household to the total number of components. '% of female workers in HH' is the ratio of employed women to all women in working age within the household. '% of Head of HH in formal sector' indicates the percentage of heads of household that are employed in the formal sector. '% of head of HH in unskilled employment' is the percentage of heads of household involved in unskilled activities such as construction, cleaning, etc. '% of head of HH in high-skilled employment' is the percentage of heads of household that are either professional or technicians. '% of HH receiving subsidies' indicates the percentage of households that receive either subsidies or any kind of social help.

In terms of education, even though the group shows clear signs of having accumulated on average more human capital than the poor, they are far from the upper class standards: by 2014 almost half of middle class heads of household had completed secondary school (or held even higher education levels) as compared to 33% of the poor and 78% of the upper class. It is worth noting that this indicator has improved for all three groups across the period. Zooming into the educational level, it is clear that university (and tertiary) education of the head of household is what distinguishes the middle from the upper class: while 33% of those who held secondary or higher education had a

university degree among the middle class (16% out of 49%), this percentage was more than half amongst the upper class (44% out of 78%).

Regarding employment of the head of household, this seems to be the most salient feature that separates the middle class from the poor. While the middle class shows rather high levels of employment almost 84%), for the poor less than 10% of the head of households are employed. This is also reflected in dependency ratios: almost 50% of individuals in middle class households are employed compared to around 20% among the poor. Furthermore, for those heads of household that are employed, some characteristics of their job are provided. In terms of formality, by 2014 around 71% of heads of households belonging to the middle class had a formal job, while this number lowered to 59% among the poor. Moreover, this indicator increased noticeably (for all three groups) during the decade evaluated. The heads of household belonging to the middle class are by and large employees (73%), while a smaller proportion is self-employed (23%) and a tiny sector owns a business (5%). In contrast, the poor show higher levels of heads as employees and the upper class almost duplicates the percentage that are business owners. Finally, approximately 17% of heads of household belonging to the middle class carried out unskilled, while 24% performed high-skilled (technical or professional) jobs. Even though this shows a sharp contrast with the poor (42% of heads of household among the poor that perform unskilled jobs) it is also true that the percentage performing high-skilled tasks is more than double for those belonging to the upper class.

Subsidies and consumption strategies seem to be the characteristics that differentiate the most the middle from the upper class. In fact, while in 2014 22.5% of middle class households received some kind of subsidy, only 8% of upper class households declared to receive one. In terms of consumption, by 2014 more than half of middle class household had to resort to installments to be able to access goods. Less than 9% of households in the upper class had to follow this strategy.

Difficulties in owning the dwelling is another salient feature. Around 60% of middle class households own their dwelling. Even though this might seem a large percentage in terms of other countries, in Argentina the access to mortgage credit is rather difficult. In contrast, on average 87% of those in the upper class own their dwellings, although it strikes that this percentage has been falling steadily as from 2004.

In terms of household access to services, all three groups seem to enjoy decent levels, although clearly the poor are worse-off. Finally, relying on a domestic employee for household chores also seems to be much more associated to the upper class.

To summarize, the middle class in Argentina, during the 2004-2014 period seems to be characterized by families with children, where the household head is usually an employed male and around 50% of them have completed secondary school or hold even higher levels of education, while only 16% hold a university degree. Their economic standing seems to be reasonable overall, but it must be noted that almost 1 every 4 families receive subsidies as of 2014 and that more than half of them need to resort to installments to access goods.

3.3 Variable selection

Once we have multidimensionally identified the Argentinean middle class and characterized its performance under the period of analysis, we are interested in performing a variable selection exercise. That is, our goal is to identify smaller subsets of variables, of cardinality $d, d \ll 19$ which are able to replicate as accurately as possible the results produced when departing from the original welfare space.

As explained in the previous section, three different variable selection procedures will be performed. On the one hand, it is interesting to explore which variables are relevant to assess multidimensional well-being. On the other hand, we would like to zoom into the poor/middle-class and identify which is the smallest subset of variables that preserve the original grouping conformation. This exercise is also done for the middle/upper-class divide. In fact, it is interesting to explore whether distinguishing the middle class from the poor is relatively easier than doing so with respect to the upper class. It is worth noting that the subset of variables selected for each of the three procedures does not necessarily need to coincide. The first exercise focuses on the variables that best describe the whole cumulative distribution of the projected welfare index. The other two exercises, instead, zoom into two particular points of that distribution, i.e., the cut-offs chosen to identify the poor, the middle class and the upper class.

All three procedures, as explained before, rely on the blinding strategy developed by Fraiman *et al.* (2008). The basic idea behind this method is to generate vectors that for a random subset of variables belonging to the original space replace the actual value for each observation with its conditional mean based on the r-nearest neighbour (r-NN) estimates. With this "blinded" vectors, each variable selection exercise performs a different test that in essence compares the results obtained with the variables in the original space and those with the "blinded" subset. The aim is to find the smallest set of variables that may reproduce fairly accurately the results obtained with the whole set of variables belonging to the original space. In essence, this method goes through the

variables in the original datasets (and their possible combinations) and leaves out variables that: *(i)* contain redundant information (i.e., are highly correlated to others); or *(ii)* only add noise.

The first variable selection procedure aims at retaining the smaller number of variables that may reproduce fairly accurately the cumulative distribution function of the unidimensional welfare index generated from the projection of the data belonging to the original space. As already explained we compare the cumulative distribution produced by both, the original variables and all possible "blinded" vectors. This is achieved by implementing hypothesis tests where the null hypothesis is that the projections do not differ, following the procedure derived in Edo *et al.* (2016). For values larger than 0.2, the subset of variables is said to represent fairly well the distribution produced by the variables in the original set. The aim is then to retain the smaller subset of variables for which the p-value of the test is larger than 0.2. Results are surprisingly stable across quarters and years (see Table A.2 in the Appendix). In fact, for *each* quarter only 4 variables seem to be relevant. This means that the unidimensional welfare index may be fairly well approximating using only 4 out of the original 19 variables. Surprisingly enough, the number of relevant variables is the same for *all* quarters, even though the variables selected in each of them are different. Nevertheless, some of them are selected in several quarters while others are never chosen. On average, the more relevant variables seem to be: the consumption strategy (selected in 95% of all quarters), income (72%), type of occupation (70%) and domestic employee (62%). Two remarks follow from these results. In the first place, it is clear that income plays a central role in determining well-being. Indeed, it is one of the most relevant variables identified. Nevertheless, it is not the only one. This result is consistent with previous literature that suggests that in spite of being very important income does not capture all of the dimensions of welfare.

The other two variable selection exercises focus on the middle-class/poor and middle/upper-class divide. The aim in this case is to find the smallest subset of variables that is able to preserve the original group allocation. The procedure follows Fraiman *et al.* (2008) and its central idea is to retain the smaller subset of variables such that at most 5% of the observations are reallocated. We will start with the middle-class/poor divide. In this case, for most quarters the smallest subset selected contains two variables (see Table A.3 in Appendix). This means that almost all individuals that are defined either as poor or middle class with the initial set of 19 variables are correctly classified based on this much smaller set of two variables. Once again, the most salient variable seems to be the consumption strategy: it is selected in 82.5% of quarters. Whether the head of household is employed is the second most selected variable (45% of quarters). It seems therefore that not being forced to buy in installments as well as having an employment are the two features

that distinguish the most the middle class from the poor. Unlike the previous exercise, income seems to be less important: only 12.5% of quarters retain this variable, standing in 5th place. It is also remarkable that for the 15 quarters (out of 43) that retain only one variable, only one of them retains income and 13 of them retain the consumption strategy.

When focusing in the middle/upper-class divide, the minimum subset that preserves the classification assigned by the original space (allowing for at most 5% of re-allocation) contains, on average, 3 variables (see Table A.4 in the Appendix). This means that one additional variable is needed to differentiate the middle from the upper class than what was needed to differentiate it from the poor. Surprisingly enough, the first two variables in terms of importance are the same: consumption strategy (selected in 86% of quarters) and whether the head of household is employed (62%). They are followed by having a domestic employee (38%) and income comes in 4th place (26%). Therefore, as in the poor/middle-class divide, we need to resort to the consumption strategy and to the head of household's occupational status to separate the middle from the upper-class. But in this case, an additional variable is needed to ensure a correct classification: having a domestic employee seems to be a key variable in differentiating the upper from the middle-class.

The three variable selection exercises carried out point to some very interesting conclusions. In the first place, the fact that in all three cases the reduced space needs to resort to more than one variable to reproduce the original welfare space is a clear indication of the multidimensionality of welfare. That is, more than one dimension is needed to identify welfare and to distinguish among groups. Secondly, income seems to be relevant in terms of defining welfare but it appears to be less useful to identify clear cut-offs that separate the middle-class from the poor and the middle from the upper class. This may be interpreted in the following way: income is certainly a key determinant of welfare, but is less reliable as a tool to distinguish among groups. Thirdly, the results reveal that is relatively easier to establish a lower bound that separates the middle class from the poor than to identify an upper threshold that distinguishes the former from the upper class. This is consistent with previous literature that points to the harder difficulties found in establishing a so-called "richness line" as compared to "poverty lines" (see Edo and Sosa Escudero, 2012)[26].

[26] A practical consideration is necessary at this point. This variable selection procedure proposed implies going over *all* possible combination of variables, for each quarter and year (more than 40 subsets), for the three exercises. This results in over 500,000 possible combinations $(2^{19} - 1)$ for each subset within each of the three exercises. It is computationally very expensive. To solve this issue, we resort to a Genetic Algorithm (GA) derived in Edo *et al.* (2016). Intuitively, the GA identifies a subset of variables that produce results fairly similar to those produced with the original space (a similar unidimensional welfare index in the first exercise, a similar classification into groups for the last two). Once this subset is chosen by the GA, then all

3.4 Compararing multidimensional to unidimensional identification approaches

A natural question that emerges from the previous analysis is to what extent the middle class identified from a unidimensional identification approach would differ from the one defined in this paper. To that end, in this section we compare the grouping classification derived from our approach to a unidimensional one. In particular, we take as a unidimensional approach a relative measure based on income, where middle class is defined as those individuals belonging to the 25^{th}-90^{th} multidimensional percentiles. This implies that those whose income falls below the 25^{th} multidimensional percentile are considered poor and those belonging to the highest decile are deemed upper-class. Table 2 shows a misclassification matrix, where we compare the grouping classification made by the unidimensional and multidimensional approaches.

Table 2. Misclassification Matrix – Conditional Probabilities

		MULTIDIMENSIONAL IDENTIFICATION		
		Poor	**Middle Class**	**Upper Class**
UNIDIMENSIONAL IDENTIFICATION	**Poor**	29.5	22.5	8.6
	Middle Class	64.3	67.1	66.4
	Upper Class	6.2	10.4	25.1
	Total	*100*	*100*	*100*

Source: own calculations based on EPH.

Notes: 'Multidimensional identification' refers to the classification obtained by applying the approach presented in this paper. 'Unidimensional identification' refers to the classification derived of a relative measure based on income that defines the middle class as those individuals earning more than the lowest 25% of the income distribution but less than the highest 90%. Each cell represents a conditional probability of belonging to a group in the unidimensional approach given a previous classification of the multidimensional approach. For instance, the cell in the second column and first row should be read as "of all individuals classified as middle class by the multidimensional approach, 22.5% would be classified as poor by a unidimensional measure".

possible combinations within this selected subset are checked with the aim of retaining the smaller one that ensures similarity of results to those produced by the original space.

Each column in Table 2 takes one of the groups (poor, middle or upper class) classified by the multidimensional approach and states how this classification would have been if a unidimensional approach would have been followed. Three important conclusions may be derived from Table 2. First of all, to a large extent both approaches coincide: 67% of individuals identified as middle class by the multidimensional approach would be classified as middle class by the unidimensional approach[27]. Nevertheless, this means that one third of the individuals would be misclassified, which strengthens our previous conclusion regarding that the complexity of welfare may not be reduced to one monetary indicator. Secondly, a relevant question is whether a unidimensional approach underestimates (or overestimates) well-being. Results show that this point is not trivial. 66% of individuals would have been classified as upper-class by the multidimensional approach and yet a unidimensional measure would have considered them middle class. This clearly indicates an underestimation of well-being. On the contrary, of all the individuals a multidimensional approach would have classified as poor, 64% would have been assigied to the middle class by a unidimensional approach. In this sense the unidimensional measure would be overestimating well-being. Therefore, it is clear that there are other relevant features that vary in a non-univocal way with income. Thirdly, the misclassification is slightly larger among the upper class than among the poor. In fact, 29.5% of those multidimensionally poor are considered poor when looking only at income, compared to 25% among the rich. This is also consistent with our previous conclusion that separating the poor from the middle class is relatively easier than distinguishing it from the upper class.

4. Concluding Remarks

The present paper contributes to the literature by presenting a new approach to multidimensionally identify the middle class, through the construction of multidimensional quantiles based on a growth direction. This approach overcomes previous method's shortcomings. In the first place, it provides a truly multidimensional identification of the middle class. Secondly, it ensures that the group identified as the middle class belongs to the central region of the multivariate distribution. In third place, it provides an exhaustive classification of all observations along a well-being index, ensuring that all observations are assigned to one of three groups: poor, middle or upper class. Last but certainly not least, it introduces a procedure to reduce the dimensionality which is able to select the

[27] Table A.5 in appendix shows the intersection between the two methods. It may be appreciated that both approaches classify in exactly the same way 53.5% of all observations. Furthermore, the table shows that the probability of being poor in one approach and upper class in the other is less than 2.5%.

most relevant variables that define multidimensional well-being and that are needed to indentify the middle class.

The case of the Argentinean middle class is analyzed under this new approach during the 2004-2014 period. We depart from 19 variables associated to different aspects of well-being: per capita family income; sources of income, property and wealth; employment and education; dwelling characteristics and having a domestic employee. We find that the Argentinean middle class seems to have fared relatively well during the years under analysis, increasing its mean income and maintaining a reasonable income share. It is worth noting, however, that this is the case for both the poor and the upper class, which suggests that this may be reflecting the path of the economy in general rather than the particular evolution of the middle class as a specific group. In terms of socio-demographic features, we find that the middle class is characterized by families with children, where the household head is usually an employed male. Around 50% of them have completed secondary school or hold even higher levels of education, while only 16% hold a university degree. Their economic standing seems to be reasonable overall, but it must be noted that almost 1 every 4 families receive subsidies as of 2014 and that more than half of them need to resort to installments to access goods. The variable selection procedure identifies 4 dimensions as being relevant to determine multidimensional welfare: the consumption strategy, per capita family income, the type of occupation as well as relying on a domestic employee. Furthermore, by zooming into the middle class/poor divide, we identify two variables as being key to preserve the grouping determined by the original space: consumption strategy and head of household's employment status. One additional variable is needed to identify the middle/upper class divide: relying on a domestic employee. This variable selection is consistent with previous findings of the literature. On the one hand, well-being is a complex phenomenon for which income is rather an imperfect proxy. Secondly, it confirms that distinguishing the middle from the upper class is far more difficult than separating that group from the poor. Finally, by comparing our multidimensional approach to a unidimensional one based solely on income we find that even though to a large extent they are consistent, one third of individuals would considered middle class by the former would be misclassified by the latter. Moreover, it is not clear whether this unidimensional measure underestimates or overestimates well-being, demonstrating that there are other relevant features that vary in a non-univocal way with income. Finally, the misclassification is slightly larger among the upper class than among the poor, confirming the previous conclusion that establishing lower thresholds is relatively easier than setting upper bounds.

The latter confirms that further research should be focused on establishing alternative multidimensional lower thresholds, probably defined in absolute terms. That is, instead of arbitrary establishing a poverty rate, it could be interesting to explore multidimensional absolute approaches to defining poverty. Combined with the proposed multivariate approach to separate the middle from the upper class, this would yield a middle class measure that could vary in terms of size while ensuring a robust identification of the upper threshold.

References

Atkinson, A. and Brandolini A.: On the identification of the Middle Class. ECINEQ Working Paper No. 217 (2011)

Banerjee, A. and Duflo, E.: What is Middle Class About the Middle Classes Around the World? MIT Department of Economics Working Paper No. 07-29 (2007)

Barro, R.: Determinants of Democracy. Journal of Political Economy 107(S6), 158-183 (1999)

Beach, C.M.: Dollars and dreams: A reduced middle class? Alternative explanations. Journal of Human Resources 24, 162-193 (1989)

Birdsall, N., Graham, C. and Pettinato, S.: Stuck in the Tunnel: Is Globalization Muddling the Middle Class? Center on Social and Economic Dynamics Working Paper No. 14 (2000)

Chaudhuri, P.: On a Geometric Notion of Quantile for Multivariate Data. Journal of the American Statistical Association. 91(434), 862-872 (1996)

Cruces, G., López-Calva, L. and Battiston, D.: Down and Out or Up and In? Polarization-Based Measures of the Middle Class for Latin America. Anales de la Asociación Argentina de Economía Política; XLIV Reunión Anual, Buenos Aires (2009)

Easterly, W.: The Middle Class Consensus and Economic Development. Journal of Economic Growth. 6(4), 317-35 (2001)

Edo, M. and Sosa Escudero, W.: Tracking the Evolution of the Middle Class in Argentina 1991-2012. Anales de la Asociación Argentina de Economía Política, XLVIII Reunión Anual, Trelew (2012)

Edo, M., Sosa Escudero, W. and Svarc, M.: Measuring the middle class in many dimensions: Argentina 2004-2014. Anales de la Asociación Argentina de Economía Política, LI Reunión Anual, San Miguel de Tucumán (2016)

Davis, J. C. and Huston, J. H.: The Shrinking Middle Class Income: A Multivariate Analysis. Eastern Economic Journal. 18 (3), 277–285 (1992)

Foster, J. and Wolfson, M.: Polarization and the decline of the Middle Class: Canada and the US. Journal of Economic Inequality 8(2), 247-273 (2009)

Fraiman, R. and Pateiro-Lpez, B.: Quantiles for finite and infinite dimensional data. Journal of Multivariate Analysis. 108, 1–14 (2012)

Gayo, M.: Revisiting middle-class politics: a multidimensional approach evidence from Spain. The Sociological Review. 61, 814-837 (2013)

Gigliarano, C. and Muliere, P.: Measuring Income Polairzation Using an enlarged Middle Class, Working Paper 271, ECINEQ (2012)

Gigliarano, C. and Mosler, K.: Measuring Middle-Class Decline in One and Many Attributes, Quaderno di Ricerca 333 (2009)

Hallin, M., Paindavaine, D. and Siman, M.: Multivariate Quantiles and Multiple-Output Regression Quantiles: From L1 Optimization to Halfspace Depth. The Annals of Statistics. 38 (2), 635-669 (2010)

Kong, L. and Mizera, I.: Quantile Tomography: Using Quantiles with Multivariate Data, Statistica Sinica. 22, 1589-1610 (2012)

Ravallion, M.: The Developing World's Bulging (but Vulnerable) 'Middle Class'. World Bank Research Working Paper 4816 (2009)

Serfling, R.: "Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization." Journal of Nonparametric Statistics, 22, 915-936 (2010)

Székely, M., Lustig, N., Cumpa, M. and Mejía, J.A.: Do we know how much poverty there is?, Working Paper 437, Inter American Development Bank (2000)

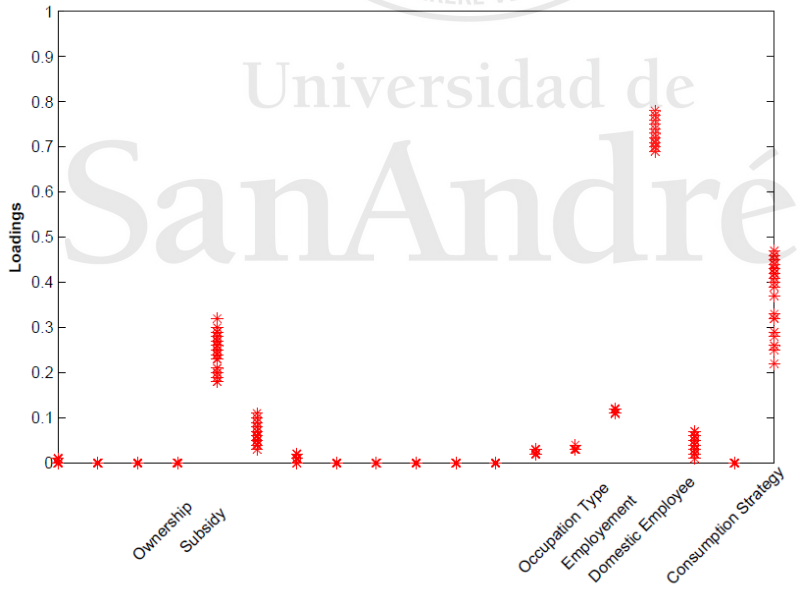
Weber, M.: *The Protestant Ethic and the Spirit of Capitalism*. Germany (1905)

Appendix

Table A.1. Household Welfare Variables[28]

Category	Variable	Description
Income	Income (amount)	Amount of Per Capita Family Income
Sources of income, property and wealth	Income from Rent	Indicates whether HH receives income from renting a property
	Income from Rent (amount)	Amount received from renting a property
	Income from Business	Indicates whether HH receives income from a business in which nobody of the HH works
	Income from Business (amount)	Amount received from a business in which HH does not participate actively
	Income from Financial Assets	Indicates whether HH receives income from financial assets
	Income from Financial Assets (amount)	Amount received from financial assets
	Ownership	Indicates whether the dwelling is owned by the HH
	Subsidy	Indicates whether HH receives subsidies
	Consumption Strategy	Indicates whether HH relies on installments to acquire good
Employment and Education	Employment	Indicates whether the head of HH is employed
	Occupation Type	Indicates the type of occupation of the head of HH (unskilled, low-skilled, technical or professional labour)
	Literacy	Indicates whether the head of HH is literate
	Education Level	Indicates the level of education of the head of HH
Dwelling Characteristics	Materials	Index that combines information on water source location and type
	Water access	Index that combines information on water source location and type
	Sewage	Index that combines information on sewage: toilet existance, location and type of discharge
	Location	Identifies whether the dwelling is located in a risky area: slum, flood zone, garbage dump zone
Other	Domestic Employee	Indicates whether HH relies on a domestic employee for HH chores

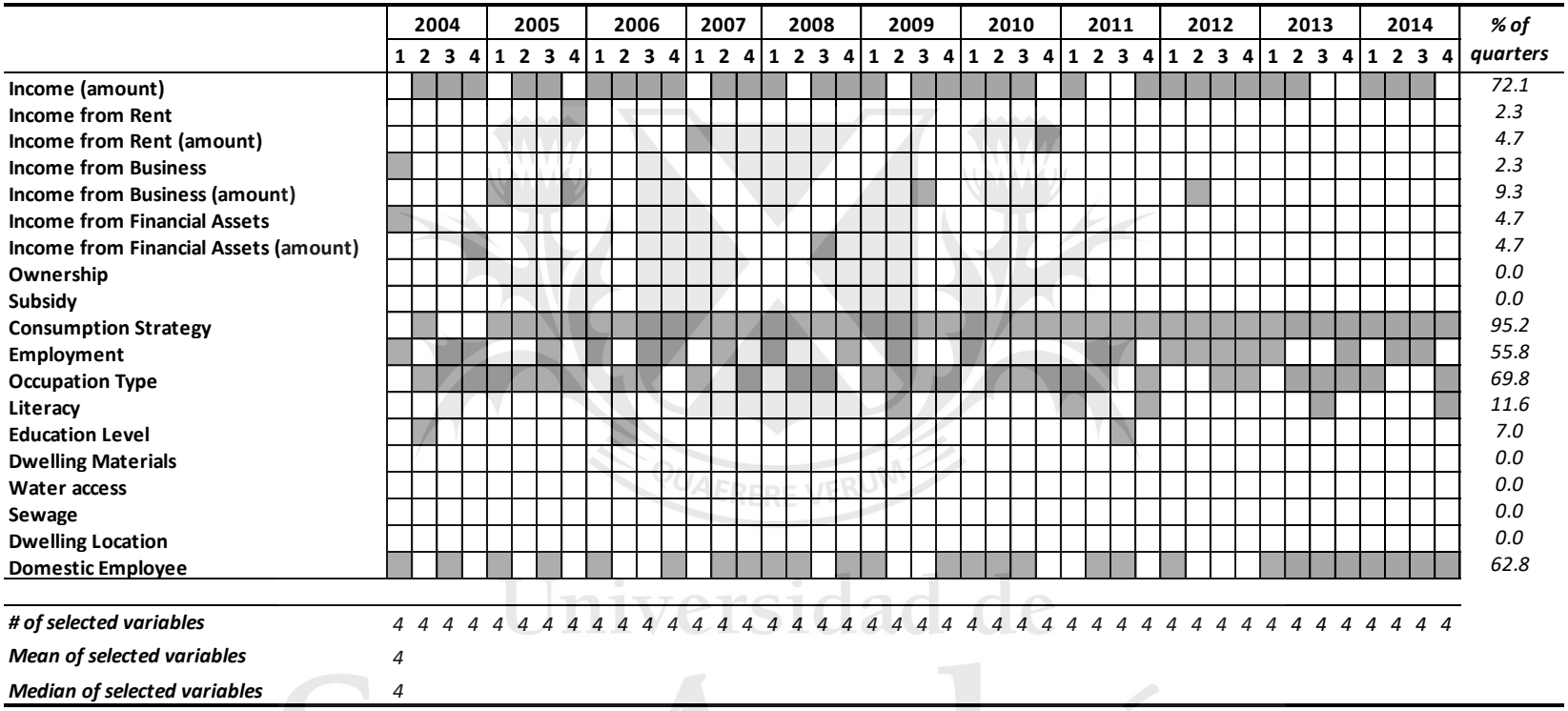
Figure A.1. First Principal Component Loadings, Argentina 2004-2014



Source: own calculations based on EPH
Notes: the X-Axis shows the 19 variables selected in the original space. Y-Axis indicates the absolute value of the weight assigned by the first principal component to each variable. The different colors represent all years and quarters.

[28] All income-related variables were deflated to make them comparable across quarters.

Table A.2. Variable Selection for Welfare Index, Argentina 2004-2014[29]

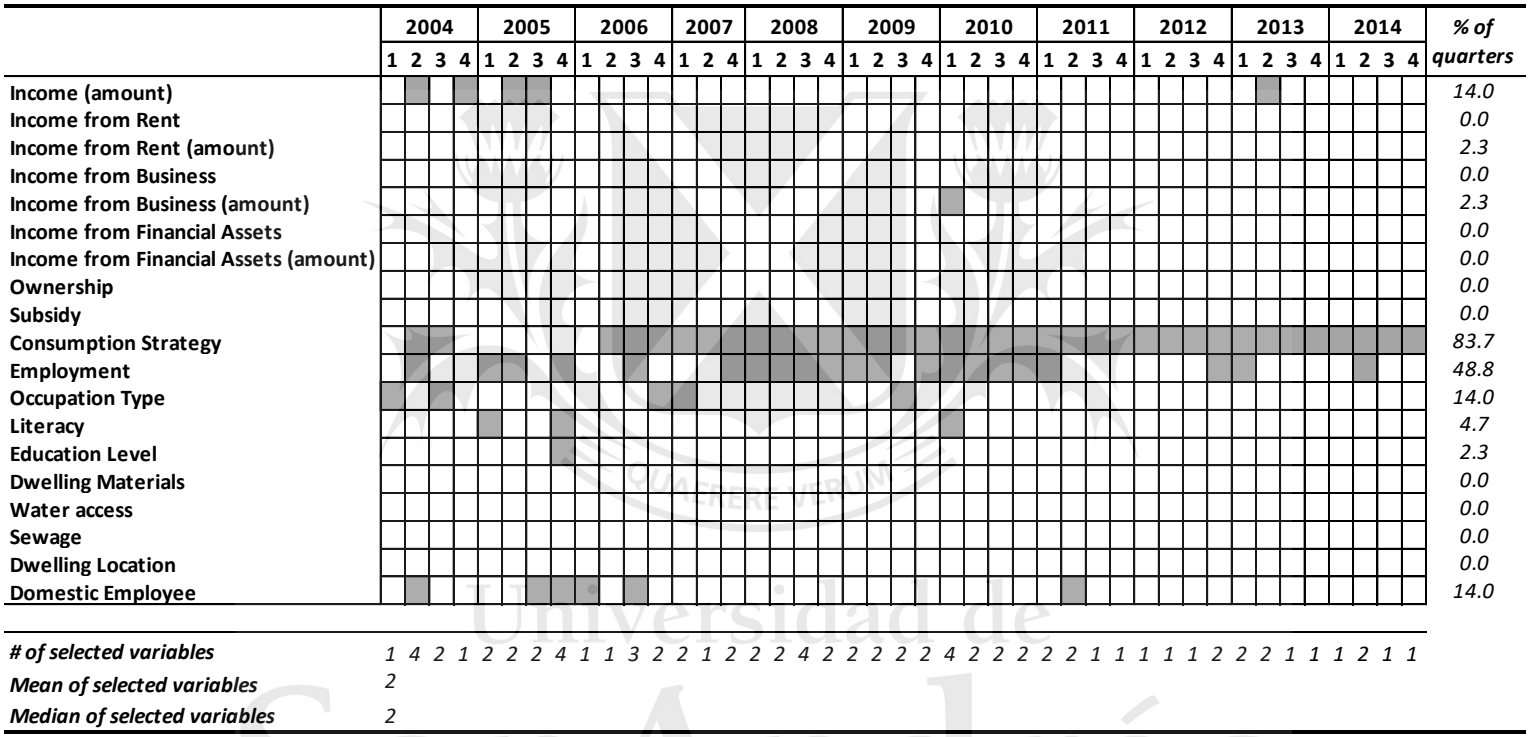


Source: own calculations based on EPH

Notes: each column shows the variables selected following the blinding strategy proposed in Fraiman *et al.* (2008) and the Edo *et al.* (2016) approach. For each quarter, the variables in gray represent the smallest subset of variables that reproduce fairly accurately the unidimensional welfare index cdf. The last row of each column shows the number of variables included in the subset for each period. Each row shows for a variable the number of periods for which it was selected. The last column of each row indicates the percentage of quarters for which the variable was selected.

[29] The EPH was not carried out for the third quarter of 2007.

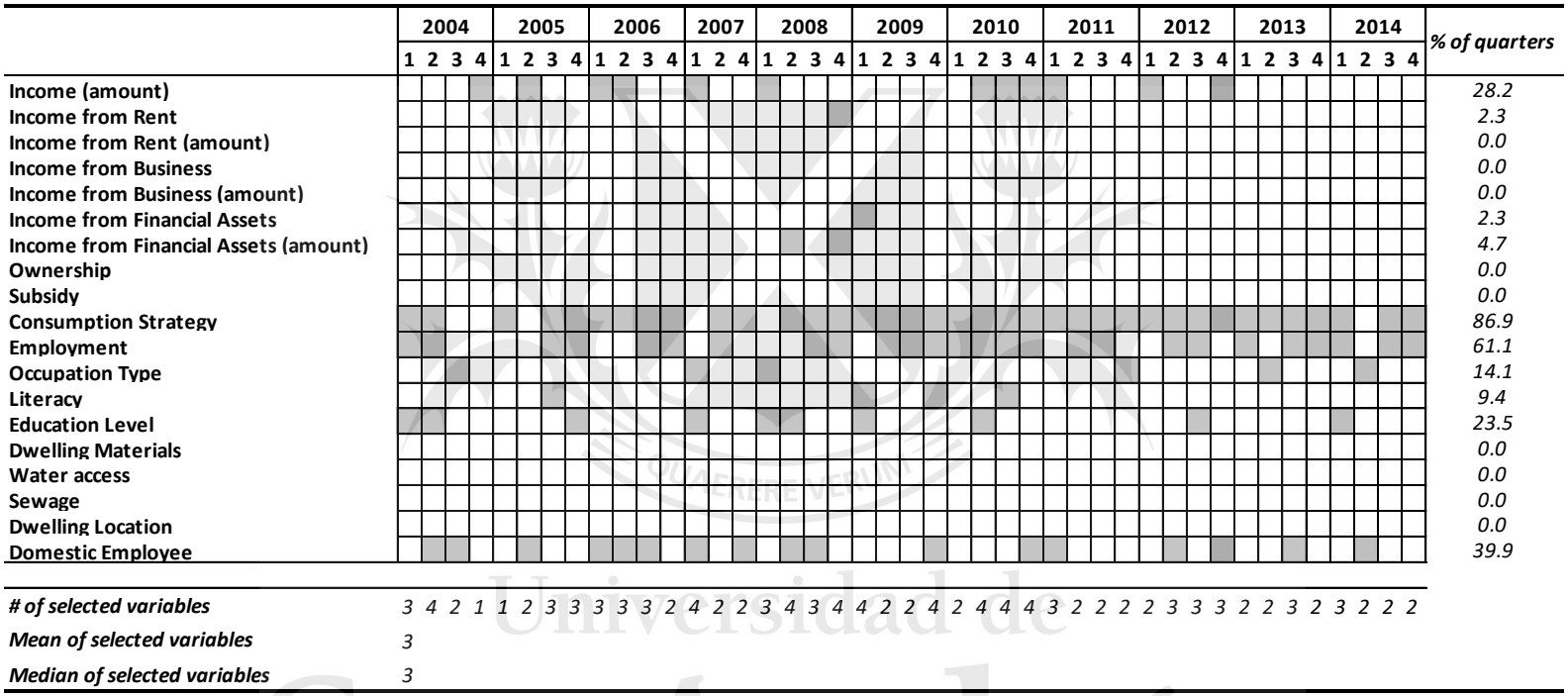
Table A.3. Variable Selection for Middle Class/Poor Divide, Argentina 2004-2014



Source: own calculations based on EPH

Notes: each column shows the variables that were selected following the blinding strategy proposed in Fraiman *et al.* (2008). For each quarter, the variables in gray represent the smallest subset of variables that preserve the original classification into middle class and poor for at least 95% of observations. The last row of each column shows the number of variables included in the subset for each period. Each row shows for a variable the number of periods for which it was selected. The last column of each row indicates the percentage of quarters for which the variable was selected.

Table A.4. Variable Selection for Middle/Upper Class Divide, Argentina 2004-2014



Source: own calculations based on EPH

Notes: each column shows the variables that were selected following the blinding strategy proposed in Fraiman *et al.* (2008). For each quarter, the variables in gray represent the smallest subset of variables that preserve the original classification into middle and upper class for at least 95% of observations. The last row of each column shows the number of variables included in the subset for each period. Each row shows for a variable the number of periods for which it was selected. The last column of each row indicates the percentage of quarters for which the variable was selected.

Table A.5. Misclassification Matrix - Intersection

		MULTIDIMENSIONAL IDENTIFICATION		
		Poor	**Middle Class**	**Upper Class**
UNIDIMENSIONAL IDENTIFICATION	**Poor**	7.4	14.6	0.9
	Middle Class	16.1	43.6	6.6
	Upper Class	1.5	6.8	2.5
	Total	*25.0*	65.0	10.0

Source: own calculations based on EPH.

Notes: 'Multidimensional identification' refers to the classification obtained by applying the approach presented in this paper. 'Unidimensional identification' refers to the classification derived of a relative measure based on income that defines the middle class as those individuals earning more than the lowest 25% of the income distribution but less than the highest 90%. Each cell represents the intersection of both methods. For instance, the cell in the second column and first row should be read as "14.6% of individuals are classified as poor by the multidimensional approach yet are considered middle class by the unidimensional approach".

CHAPTER III. Compulsory education laws or incentives from CCT programs? Explaining the rise in secondary school attendance rate in Argentina[*]

Abstract

Argentina has traditionally stood out in terms of educational outcomes among its Latin American counterparts. Schooling of older children, however, still shows room for improvement especially among the more vulnerable. Fortunately, during the last years a sizeable improvement in attendance rates for children aged 15 through 17 took place. This could be related to the 2006 National Education Law that made upper-secondary education compulsory. In this paper, instead, we claim that the Asignación Universal por Hijo (Universal Child Allowance, AUH) -a massive conditional cash transfer program implemented in 2009 in Argentina- may be mostly responsible for this improvement. Using a difference-in-difference strategy we estimate that the program accounts for a 3.9 percentage point increase in the probability of attending secondary school among eligible children aged 15 through 17. The impact seems to be led by boys and is more relevant for children living in larger families where the head of household has a lower educational level.

*JEL Code***:** I24, I25, I38

Keywords: conditional cash transfers, education, attendance, Argentina.

[*] This is joint work with Mariana Marchionni and Santiago Garganta from CEDLAS-Universidad Nacional de La Plata and CONICET.

1. Introduction

Argentina has traditionally stood out within Latin America in terms of education. Since the very creation of the National Education System in 1884, primary education has been mandatory in Argentina. This and the free public provision of educational services have allowed to reach almost perfect rates of primary school attendance, which have remained relatively stable above 97% since the 1980s and are comparable and even higher than those of developed countries (Marchionni and Alejo, 2015).

In contrast, secondary education has not always been mandatory in Argentina. By the early 1990s, only the seven years of primary education were compulsory. In 1993, the Federal Education Law 24,195 (*Ley Federal de Educación*) increased compulsory education from 7 to 10 years, thus including the first stage of secondary education. The National Education Law 26,206 (*Ley Nacional de Educación*) passed in December 2006 extended compulsory education by three more years, making mandatory also the upper-secondary education level.

Secondary education indicators improved markedly since the mid-1990s, and some argue that these improvements are a consequence of the successive expansions of compulsory education (DiNIECE, 2011). For the case of the 1993 Federal Education Law, Alzúa *et al.* (2015) find a positive effect on school enrollment and attainment, but the mechanisms remain unclear since the 1993 reform combined an expansion in compulsory education with deep institutional and curricular modifications, among other changes.

Over the last decade, the net school attendance rate for the group aged 15 to 17 – the upper-school age range – rose by almost 4 percentage points, from 82.9% in 2004 to 86.6% in 2014[31]. Our first hypothesis is that this improvement was not caused by the 2006 law. First, neither the law nor accompanying policies had enforcement mechanisms embedded in their design. Therefore, it is unclear through which channels the law may have affected school attendance. Second, three years after the law was passed, attendance rates for the group aged 15 to17 remained virtually unchanged. Only since 2010 school attendance for individuals in this age group started to show clear signs of growth.

But if the 2006 National Education Law showed no impact on net attendance of those aged 15-17, what is driving the increase in those rates as of 2010? What is bringing children aged 15 to 17 –

[31] Net school attendance rate is the percentage of children in a given age group that attend the educational level that officially corresponds to that age (UNESCO).

especially those most poor – to stay in school? In this paper we claim that the *Asignación Universal por Hijo* (Universal Child Allowance, AUH) a program implemented in Argentina in late 2009 may be driving this increase in attendance rates.

The AUH is a massive conditional cash transfer program (CCT) targeted at children under 18 years old living in poor families with no registered workers in the formal employment sector. The program currently benefits 29% of all children and approximately 15% of total households in the country (ANSES, 2014). As any typical CCT program, the reception of the transfer is conditional on complying with children's health requirements and school attendance at all compulsory levels. The high economic incentives introduced by the AUH and their conditionalities may probably reduce the probability of dropping out of secondary school compared to the counterfactual situation in absence of the program.

Estimating the causal effect of the AUH on school attendance, however, represents a rough task. The AUH was not assigned randomly nor was it accompanied by a publicly available comprehensive dataset that allows for assessing the program. We thus resort to the Permanent National Household Survey (*Encuesta Permanente de Hogares,* EPH*)* carried out in Argentina. We classify children in upper-secondary age-range (15 to 17) as potential beneficiaries according to whether their parents comply with the program's eligibility requirements. We compare the probability of secondary school attendance of both groups (eligible and not eligible) over time following a difference-in-difference approach.

Our estimates suggest that the AUH increased the probability of attending secondary school among eligible children aged 15-17 by 3.9 percentage points. The impact seems to be led by boys and is more relevant for children living in larger families where the head of household has lower education levels. The effect on younger children is statistically significant yet very small: 0.4 percentage points for those in primary school age range (6 through 11) and 0.8 percentage points for those in lower secondary (12 through 14 years old). The results hold across different specifications and robustness analysis.

This paper intends to make contributions in two ways. First of all, it adds to the literature on the impact of CCT programs on educational outcomes. Secondly, it provides evidence of the effects of the *Asignación Universal por Hijo*, thus generating input for future improvements of the program. Finally, this work also seeks to highlight the fact that compulsory education laws by themselves are not enough to affect schooling.

The rest of the paper is organized as follows. Section 2 expands on compulsory education legislation in Argentina while presenting evidence on the evolution of net attendance rates over the last decade. Section 3 describes the AUH and discusses the channels that may affect schooling decisions. Section 4 presents the data and methodology. Section 5 and 6 discuss the main results while Section 7 concludes and points to further research.

2. Compulsory education laws and school attendance in Argentina

Compulsory education laws are motivated by the potential social benefits and positive externalities coming from an expansion of the overall education attainment which promotes economic development (Oreopoulos, 2006a). These laws may affect attendance rates through different channels. In the first place, the human capital model of school choice perceives education as an investment (Becker, 1975) and hence depends on intertemporal benefits and costs of schooling. Consequently, compulsory education may prevent a probably optimal decision of leaving school. However, compulsory attendance laws may rise lifetime welfare if they generate positive externalities or under the presence of suboptimal school attainment (Oreopoulos, 2006a; Eckstein and Zilcha, 1994), which is likely among the more vulnerable children in developing countries like Argentina. Secondly, these legislations may trigger implicit enforcement mechanisms, by imposing social stigma to those who fail to comply with the rule. Fulfillment of mandatory schooling may also affect future opportunities in the labor market if, for instance, legal educational requirements are set as a condition to enter the formal employment sector (Alzúa *et al*., 2015). Finally, other public policies accompanying the launch of these legislations may have an impact on attendance rates by affecting the direct costs of education (abolition of tuition fees), the quality of education (increase in educational budget, drastic changes in the curricula) or the availability of nearby educational facilities (large-scale infrastructure programs), among others.

Unfortunately, evidence of the impact on attendance rates of changes in compulsory education laws is relatively scarce. Most studies concentrate on the effects regarding labor market outcomes (Angrist and Kruger, 1991; Acemoglu and Angrist, 2000; Oreopoulos, 2006a and 2006b). Even though some studies document the improvement of attendance rates following mandatory education laws (Goldin and Katz, 2008; Lleras Muney, 2002; Oreopoulos, 2006a), the mechanism through which the effect operates is not entirely clear. Compulsory education laws are usually launched together with other policies aiming at increasing school attendance. Therefore, some or all of the abovementioned channels operate at the same time, hindering the possibility of isolating the impact of the expansion of compulsory education by itself.

Regarding Argentina, while primary education has always been mandatory, it was only in the early 1990s that compulsory schooling expanded to secondary education. The Federal Education Law, passed in 1993, increased mandatory education from 7 to 10 years of schooling, thus including the first stage of secondary education (children up to 14 years old). Later, in 2006, the National Education Law added three more years of compulsory education, covering also the upper-secondary level (youths between 15 and 17 years old)[32]. Table 1 summarizes the timing and scope of these reforms.

Table 1. Extension of compulsory education in Argentina

Age	Common Education Law Year: 1884	Federal Education Law Year: 1993	National Education Law Year: 2006	Modification to National Education Law Year: 2015
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
Compulsory Years	7	10	13	14

Indicates compulsory age

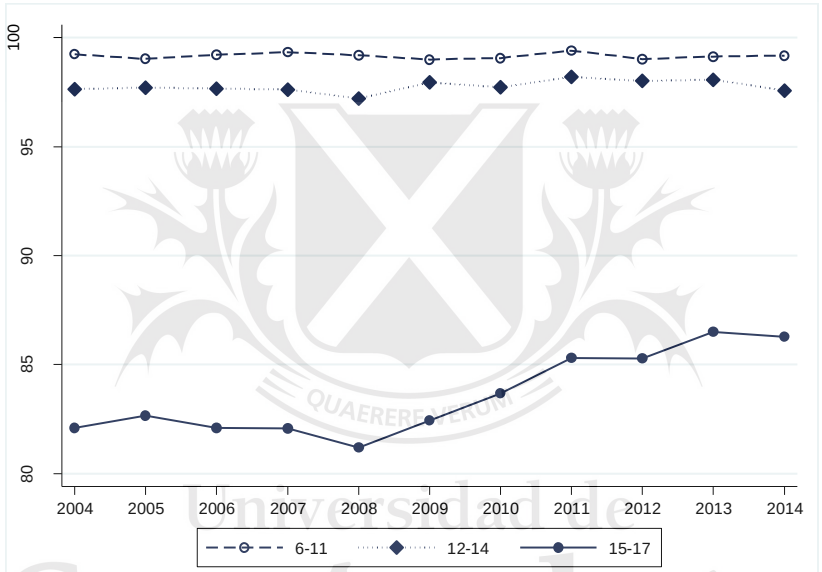
Sources: Common Education Law (1884), Federal Education Law (1993), National Education Law (2006).

Some argue that these successive expansions in mandatory schooling are responsible for the observed improvements in secondary education indicators since the early 1990s in Argentina

[32] Only four other Latin American countries have passed equivalent legislation (i.e. mandatory schooling for both primary and secondary education): Uruguay in 2008, Chile and Brazil in 2009 and Mexico in 2013 (Ruiz and Schoo, 2014).

(DiNIECE, 2011). However, the evidence is not so clear. Alzúa *et al*. (2015) evaluate the impact of the 1993 law by taking advantage of the different timing in the implementation of the reform. They find that the 1993 law was followed by a notable increase in gross enrollment rates and had a positive impact on years of schooling for children aged 13-14. However, as stated by the authors, the main mechanism driving the effect is hard to identify since the new legislation was accompanied by changes in the curricula and a strong expansion of the education budget to finance investment in school infrastructure as well as teacher's training.

Figure 1. Net school attendance rates by age group



Source: own estimations based on EPH.

Notes: Net school attendance rate is the percentage of children in a given age group that attend the educational level that officially corresponds to that age (UNESCO). Ages 6-11 correspond to primary school; ages 12-14 and 15-17 correspond to lower and upper secondary school, respectively.

Figure 1 shows that by 2004, net attendance rates for children aged 6 to 11 (primary school age) and 12 to 14 (lower-secondary school age) were above 97% and remained rather stable over the following decade. Compared to these younger children, those aged 15 to 17 exhibit markedly lower attendance rates (82% in 2004). Even though for this latter group education became compulsory in

2006, net attendance rates remained mostly unchanged over the following three years[33]. Only after 2009 net attendance rates started to significantly grow for 15-17 year-olds, from 82.9% in 2009 to 86.6% in 2014, i.e. an almost 4-percentage-point increase[34].

The preliminary evidence in Figure 1 suggests that the 2006 National Education Law had no impact on net attendance rates on the first three years after its implementation, which is not surprising given that there were no companion measures that could have encouraged school attendance. In fact, even though there was a large expansion of the educational budget, new funds were almost entirely absorbed by salaries, with no investment in training or systematic infrastructure development, and only quite limited changes in the curricula[35]. Moreover, despite some specific programs were developed to complement the new education law they were more focused on establishing an adequate normative framework and on improving institutional arrangements than in providing direct or indirect incentives to school attendance (UNICEF, 2012)[36].

But if the 2006 National Education Law had no impact on attendance rates for those aged 15-17 three years after its implementation, what is driving the increase as of 2010 shown in Figure 1? In this paper we claim that the *Asignación Universal por Hijo* (Universal Child Allowance, AUH) program implemented in Argentina in late 2009 is responsible for encouraging children aged 15 through 17, especially poor children, to stay at (or return to) school. In fact, Figure 2 shows that the improvement of upper-secondary net attendance rates after 2009 was driven by the most vulnerable children, i.e. the target group of the AUH. Net attendance rates for youths aged 15 to 17 in the first quintile of the income distribution increased 8 percentage points in the last decade: almost 3 percentage points between 2004 and 2009 (from 72.8% to 74.6%) but more than 5 percentage points between 2009 and 2014 (from 75% to 80.5%). Net attendance rates for those in the top quintiles have remained mostly unchanged over the last decade.

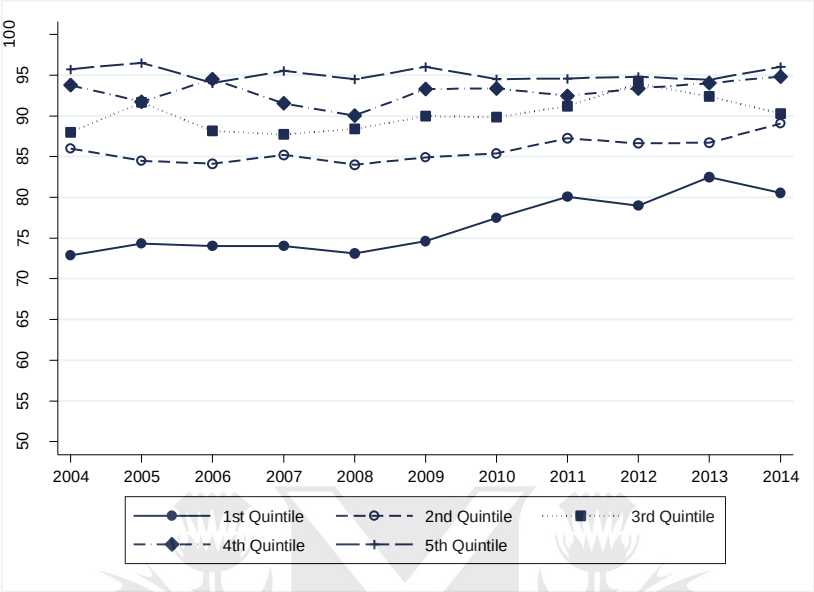
[33] In fact, attendance rates for the group of 15-17 year-olds follow a similar pattern to the 12-14 year-old group over the 2004-2009 period, even though the latter group was not affected by the law. This is confirmed by a difference-in-difference estimation. These results are available upon request.

[34] Administrative data shows a very similar pattern for secondary school enrollment (DiNIECE: http://portales.educacion.gov.ar/diniece/). However, administrative data is not available by age group.

[35] The 2005 Education Funding Law 24,075 (*Ley de Financiamiento Educativo*) introduced a gradual expansion of the educational budget, with the aim of reaching 6 percent of GDP by 2010. This implied an increase in per-student expenditure in Argentina, but the country lacked improvements in terms of the efficiency of this investment, in particular the pedagogical and organizational transformations to facilitate the improvement of education results (Auguste, 2012).

[36] For instance, the *Plan Nacional de Educación Secundaria Obligatoria* (2009-2011 and 2012-2016).

Figure 2. Net attendance rates for 15-17 year olds by income quintile.



Source: own estimations based on EPH.

Note: quintiles of the distribution of per capita family income.

3. The AUH program and the incentives to school attendance

The AUH was launched in November 2009 and represents a massive conditional cash transfer (CCT) program that focuses on children under 18 years old living in poor and informal households. It was designed to extend the social protection network in Argentina, which used to be tied to the formal employment sector, to the more vulnerable groups of the population. The magnitude of the benefit as well as the expansion in the number of beneficiaries have no precedents in the Argentinian social policy, formerly characterized by small scale and targeted programs.

The AUH awards a monetary transfer to households with children where neither parent is registered in the formal sector. This includes inactive, unemployed or informal workers earning less than the minimum wage.[37] Each beneficiary household can perceive a transfer per child under 18 years old up to a maximum of five dependent children.[38] Currently, more than 3.5 million

[37] It is important to note, however, that monitoring this condition is not feasible in practice. This implies that informal workers earning more than the minimum wage could become beneficiaries. Nevertheless, as shown later on, both quantitative and qualitative evidence suggest that these situations are scarce, probably due to social responsibility or stigma.

[38] Transfers for disabled children have no age limit.

children and youths benefit from this program, representing almost 29% of all individuals under 18 years old and approximately 15% of total households in the country (ANSES, 2014). Regarding its budget, the AUH is one of the largest CCT programs in Latin America, with resources representing almost 0.8% of the country's GDP (Stampini and Tornarolli, 2013).

CCT programs may impact on school enrollment and attendance by relaxing family's budget constraints but also through the conditionalities they impose. As education may be regarded as a normal good its consumption could increase with household income. The conditionalities set an additional incentive to bias this increase in consumption towards investment in education. In particular, the AUH imposes sanitary and educational conditionalities in terms of periodical health controls and vaccination for children under 5 and pregnant women, and school attendance at all compulsory levels from ages 5 through 18. For this purpose, the program sets a particular payment mechanism: 80% of the subsidy is automatically received by beneficiary families on a monthly basis, and the remaining 20% is paid annually, once compliance with the conditionalities is proven[39]. If the conditions are not met, not only the 20% is not perceived but also the beneficiary is suspended from future participation in the program.

The amount of the AUH transfer has been modified several times to cope with inflation[40]. As of June 2014, the monthly transfer for each child -i.e. 80% of the total transfer- was ARS 515 which represented almost 15% of the minimum legal wage in Argentina. For a typical poor family with three children, this implied an almost 30% increase of total monthly family income. The 20% remaining amounted to ARS 1,400 per child per year, *i.e.* 62% of total family income for the same typical family and almost 100% of the minimum legal wage. Besides, noncompliance with the conditionalities leads to suspension from the program, implying the loss of the future transfers until the child turns 18. Since the AUH was launched as a permanent program with a wide support of all political parties, the transfers should be perceived as permanent income and the expected present value of the transfers should be large, thus reinforcing the commitment of beneficiaries with conditionalities.

[39] Concerning the condition on school attendance, the program originally required that the child must be enrolled in a public school. This clause, however, was never made effective given the large public opposition that claimed for a considerable fraction of vulnerable children who attends publicly subsidized private schools. In fact, 16% of all primary school students belonging to the first two quintiles of the equivalized income distribution were attending a private school in 2010. The corresponding figure for secondary school students is 14% (SEDLAC, 2015).

[40] The nominal monthly benefit per child, initially set at ARS $180, has increased on average more than 20% per year and hence its real value has remained relatively constant since 2009 (Garganta *et al.*, 2014).

According to the literature on the impact of CCT programs, the effects on 'access to school' indicators such as enrollment and attendance are usually positive (Fiszbein *et al.*, 2009; Cecchini, 2014), even though the size of the effect varies with other factors: it is larger for groups with low attendance rates, among the most vulnerable families and in programs with more generous transfers (Saavedra and García, 2012). Besides these general findings, some particular results are worth noticing. Typically, the size of the effect is larger in the secondary school level than in the primary level. For instance, both the *Oportunidades* program in Mexico (formerly known as *PROGRESA*) and *Familias en Acción* in Colombia significantly contributed to increase attendance rates, especially among secondary school children (Attanasio *et al.*, 2008; Schulz, 2004; De Brauw and Hoddinott, 2008). Also, even when focusing on secondary education, the size of the effect exhibits considerable variation: from a 2-percentage-point increase in the case of *Ingreso Ciudadano* in Uruguay to a 12-percentage-point increase in the case of *Oportunidades* in Mexico and *Bolsa Escola* in Brazil (Saavedra and García, 2012)[41]. In summary, even though the impact differs across programs and population groups, in general CCT programs improve the so-called 'intermediate objectives': better access to school, higher enrollment rates and higher attendance (Cecchini, 2014; Bastagli, 2008).

Given this evidence and the importance of the AUH – both in terms of coverage and generosity of the benefits – it is likely that it contributed to the improvement of attendance rates documented in section 2, which took place precisely after the program's inception in late 2009. Evidence of the impact of the AUH on education results is still scarce. Among a large set of well-being indicators, Paz and Golovanevsky (2014) find large and positive effects in attendance rates – around 7 percentage points – of the AUH for eligible children aged 13-17 when comparing the years 2009 and 2010 through a difference-in-difference methodology. In a recent working paper based on aggregate data from administrative sources, Cigliutti *et al.* (2015) find that secondary gross enrollment rates in Argentina rose by 2.25 percentage points due to the AUH compared to a synthetic control that consists of a linear combination of other Latin American countries[42]. The present work provides new evidence regarding the impact of the AUH on eligible's secondary school attendance. By using micro-data and following a difference-in-difference approach we extend the period of analysis to cover 6 years before and 5 years after the AUH implementation. Furthermore, we zoom into the group aged 15-17 which allows for relating our findings to the

[41] Additional evidence from the Mexican *PROGRESA/Oportunidades*, the oldest and most studied program in Latin America, shows also a significant reduction in drop-outs (SEDESOL, 2008), a fall of the gender gap in secondary enrollment (Parker, 2003) and an increase in indigenous children attendance (Escobar and De la Rocha, 2008).

[42] D'Elia *et al.* (2014) provide evidence of the AUH impact on education quality indicators.

extension of mandatory schooling while deepening the analysis of the nature of the effect by exploring heterogeneities across different sub-groups.

4. Data and empirical strategy

The AUH was neither randomly assigned nor accompanied by a publicly available comprehensive dataset that may allow for follow-ups of the beneficiary population. The absence of these features greatly determines both the data and the empirical strategy for assessing the program's impact on any outcome.

We use microdata from the Permanent National Household Survey (EPH) carried out by the Argentinian national statistical office (INDEC). The EPH gathers data on demographic, education, income and employment issues and covers 31 large urban conglomerates, representing 62% of the total population of the country. We focus on the 2004-2014 decade. The pre-intervention period (before) includes years 2004 through 2009 – the AUH was launched in November 2009 – while the post-intervention period (after) covers years 2010 through 2014.

Our sample includes children aged 15-17, *i.e*. in the upper secondary age range. Since the EPH does not include information to identify AUH beneficiaries, we aim at determining if the child is a potential beneficiary of the program by checking whether he/she meets the AUH eligibility criteria – intention to treat. Particularly, we define the 'treatment' and 'control' groups based on children's eligibility according to their parents' labor status. A child is classified as belonging to the treatment group whenever his/her parents are either inactive, unemployed, informal or self-employed workers. Because of a special regulation, children whose parents are registered employees working in the domestic service are also eligible for the AUH and hence are included in the treatment group[43]. As for the control group, it includes all children aged 15-17 for whom at least one of their parents is employed in the formal sector.

As an additional requirement for eligibility, the AUH imposes that earnings are below the minimum legal wage. Even though this condition is not verifiable for informal workers, qualitative and quantitative evidence suggests that middle and high-income informal workers opt out of the program due to social responsibility and stigma, and hence the inclusion error is small[44]. Therefore,

[43] Special Social Security Scheme for Domestic Service Employees (Law 25,239, Title XVIII).

[44] From the experience of public officials in charge of the registration to the AUH, non-poor individuals –yet not belonging to the formal sector- tend to opt out of the program either by not even starting the procedure or by not complaining when they are suspended from the benefit following audits (Pautassi *et al.*, 2013). Evidence from the last National Consumption Survey (ENGHo 2012) points in the same direction: very few

we further restrict the sample to only include children from poor households, defined as those in the first four deciles of the per capita income distribution[45].

In order to estimate the intention-to-treat impact of the AUH on secondary school attendance of eligible children we follow a difference-in-difference methodology by comparing the differences in the probability of secondary school attendance of the treatment and control groups, before and after the inception of the program. The identification assumptions are that secondary attendance rates of treatment and control groups would have evolved similarly in the absence of the program and that there was no other contemporaneous event to the implementation of the AUH that could have caused differences in the evolution of school attendance between the treatment and control groups. The latter does not appear to be a strong assumption considering no major initiatives affecting educational outcomes took place in 2009 (infrastructure expansion, teacher's training, school meals, etc.). Regarding the first assumption, it cannot be proven but we provide evidence in its favor in the next section.

As for the difference-in-difference model, we use the standard linear specification in equation (1).

$$Attends_i = \propto + \beta_1 Treat_i + \beta_2 After_i + \gamma(Treat_i . After_i) + \theta\, X_i + u_i \quad (1)$$

The output variable *Attends* is a binary indicator that takes the value 1 for children attending secondary school and 0 otherwise[46]; *Treat* is an indicator variable for the treatment group; *After* tags years after the AUH implementation (2010-2014), and X includes a set of child and household level controls (child's gender, age and squared age; head of household's gender, age, squared age, educational level and employment status) as well as other household characteristics (household size, per capita income, single parent household, female headed household, number of children under 18). We also control for time (year and quarter) and regional fixed effects, as well as for regional trends[47]. If the unobserved characteristics that remain after adding all these controls do not have a differential impact on attendance between both groups before and after the implementation of the AUH, we may claim that the γ parameter represents the causal effect of the program (Angrist and Pischke, 2009).

children belonging to the upper income deciles – less than 2% in the two top deciles – receive benefits from the AUH (Gasparini and Cruces, 2015).

[45] Results are robust to other income measures as well as other cut-offs.

[46] Unfortunately, even though the EPH includes information on the education level being attended, it does not inform the specific school-year.

[47] We use data for the first semester of each year by combining EPH's samples from the first two quarters and control for quarter fixed effects.

5. Results

Table 2 shows average net attendance rates for treatment and control groups before and after the inception of the AUH. Even though attendance rose for both groups, the increase was considerably larger among eligible children: 5.1 percentage points as compared to 1.9 for the control group. This preliminary unconditional evidence suggests that the AUH may have had the effect of rising secondary school attendance of eligible children aged 15-17 by 3.2 percentage points.

Table 2. Net Secondary School Attendance Rates

Children between 15 and 17 years old

	Treatment (i)	Control (ii)	(i)-(ii)
Before AUH	75.1	87.0	-11.9
After AUH	80.2	88.9	-8.7
Difference (After-Before)	***5.1***	***1.9***	***3.2***

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: Treatment Group includes children whose parents are either inactive, unemployed, informal or self-employed workers (or are registered employees working in the domestic service). *Control Group* includes all children aged 15-17 for whom at least one of their parents is employed in the formal sector. *Before AUH* includes years 2004-2009 while *After AUH* includes years 2010-2014.

It is worth noting, however, that given the very nature of the program – non-random assignment –, treatment and control groups differ by construction. Table 3 shows that even though the two groups share on average some features (gender, age, household's size), potential AUH beneficiaries belong to poorer households and exhibit a larger proportion of single-parent and female headed households where the head of household has lower educational attainment and is more likely to be unemployed, both in pre and post-intervention periods.

Table 3. Descriptive statistics

Children between 15 and 17 years old

	Variables	Before				After			
		Treatment Group	Control Group	Difference	P-Value	Treatment Group	Control Group	Difference	P-Value
Child	**Male**	51.5	51.0	0.5	0.6	50.1	51.4	-1.3	0.9
	Age	15.9	15.9	0.0	0.5	16.0	15.9	0.1	0.0
Head of HH	**Single-Parent**	34.7	14.4	20.3	0.0	36.9	16.6	20.3	0.0
	Female	36.6	18.2	18.4	0.0	42.0	22.7	19.3	0.0
	Age	46.4	45.4	1.0	0.0	46.1	45.4	0.7	0.0
	Years of Education	7.9	9.0	-1.1	0.0	8.4	9.5	-1.1	0.0
	Employed	73.5	89.9	-16.4	0.0	71.3	89.4	-18.1	0.0
HH	**Household Size**	5.8	5.8	0.0	0.4	5.7	5.7	0.0	0.1
	Number of Children	3.2	3.1	0.1	0.0	3.1	3.0	0.1	0.0
	Per Capita Income	184.3	285.4	-101.1	0.0	741.7	1012.5	-270.8	0.0
	Observations	***12,466***	***6,363***			***10,002***	***6,171***		

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

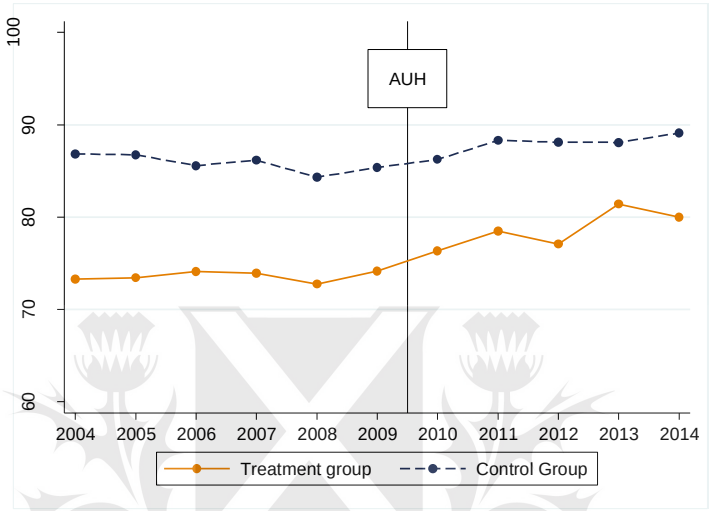
Source: own estimations based on *Encuesta Permanente de Hogares*.

Note: Treatment Group includes children whose parents are either inactive, unemployed, informal or self-employed workers (or are registered employees working in the domestic service). *Control Group* includes all children aged 15-17 for whom at least one of their parents is employed in the formal sector. *Before AUH* includes years 2004-2009 while *After AUH* includes years 2010-2014. *Number of Children* is the total number of children under 18 living in the household. *HH* stands for household.

In fact, as Figure 3 shows, treatment and control groups differ in their school attendance rates prior to the program, which is in part due to those differences in characteristics. Nevertheless, albeit attendance rates *levels* differ before the inception of the AUH, the *time patterns* are similar. This is confirmed by a pre-program common trends test: we do not find enough evidence to reject the null hypothesis that the pre-treatment trends were equal, thus reinforcing the confidence in our

identification assumption[48]. However, since 2010 just after the AUH implementation, the school attendance gap between groups started to shrink because the attendance rate of eligible children grew faster than that of the control group.

Figure 3. Net attendance rates for 15-17 year olds. Treatment and control groups.



Source: own estimations based on *Encuesta Permanente de Hogares*.

Note: Treatment Group includes children whose parents are either inactive, unemployed, informal or self-employed workers (or are registered employees working in the domestic service). *Control Group* includes children for whom at least one of their parents is employed in the formal sector. Children in both groups are aged 15 through 17 and belong to the first four deciles of the per capita family income distribution.

We now assess whether this result holds in a multivariate difference-in-difference framework and is robust to several types of controls. Table 4 shows the results of estimating the linear model of school attendance in equation 1. Models 1, 2 and 3 in the table progressively control for child´s and head of household's characteristics (child's gender, age and squared age; head of household's gender, age, squared age, educational level, employment status), other household features (household size, per capita income, single-parent household, female headed household, number of children under 18), region and time fixed effects (year and quarter), as well as regional trends. The

[48] We run a model of our outcome of interest (attendance) on a constant, the treatment dummy, year dummies and the interactions between these latter variables including only pre-intervention years. We then apply an *F* test in which the null hypothesis (*Ho*) states that all the coefficients for the interaction terms are jointly equal to zero. We find no evidence to reject the null: *Ho*: $F(5, 18{,}817)=0.47$, Prob>F=0.80. We then run a new model that includes both pre and post-program years. The null hypothesis is now easily rejected: *Ho*: $F(10, 34{,}980)=2.19$, Prob>F=0.015.

coefficient of the interaction term is positive and statistically significant across specifications, suggesting a positive impact of the AUH on school attendance of eligible children aged 15-17 of almost 4 percentage points (3.9 p.p.).

Table 4. Probability of attending secondary school

Children between 15 and 17 years old

	(1)	(2)	(3)
Treatment*After	***0.0320****	***0.0392****	***0.0388****
	(0.00817)	***(0.00890)***	***(0.00885)***
Treatment	*-0.119****	*-0.0771****	*-0.0757****
	(0.00728)	*(0.00623)*	*(0.00622)*
After	*0.0195****	*0.000711*	*0.0309*
	(0.00655)	*(0.00700)*	*(0.0433)*
Child and HH head's characteristics	Yes	Yes	Yes
Other HH Characteristics	No	Yes	Yes
Regional and Time Dummies, Regional Trends	No	No	Yes
Observations	35,002	35,002	35,002

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: OLS estimations. Dependent binary variable: *Attends*, equals 1 if the child is 15-17 years old and attends secondary level; *Treatment* equals 1 for eligible children and 0 for non-eligible children; *After* equals 1 in the period 2010-2014 and 0 for the period 2004-2009; child´s and/or head of household´s characteristics (child's gender, age and squared age, head of household's gender, age, squared age, educational level and employment status), other household characteristics (household size, per capita income, single parent household, female headed household, number of children under 18), region fixed effects (6 regions), time fixed effects (year and quarter) and regional time trends.

The size of the effect is certainly non-trivial. According to our estimates, the 3.9 percentage-point impact in secondary school net attendance implies that the AUH helped around 20,000 eligible

children aged 15-17 to stay at secondary school over the period 2010-2014. In terms of education gaps, it represents a 20% closure of the net attendance rate gap between the treatment group and those belonging to the richest quintile. Moreover, compared to other Latin American CCT programs, the impact we find for the AUH is between the 2-percentage-point effects of the Brazilian *Bolsa Escola* and the Uruguayan *Ingreso Ciudadano*, and the 12-percentage-point effects of *Familias en Acción* in Colombia and *Oportunidades* in Mexico (Saavedra and García, 2012)[49].

Placebo experiments

We perform a series of false experiments or placebo exercises to gain more confidence in the validity of the identification assumption. In this regard, we run the same linear model using only pre-treatment observations and pretending that the program took place in any year previous to 2009 – the actual implementation date of the AUH. Table 5 shows the results for five alternative fake dates: 2004, 2005, 2006, 2007 and 2008. In all cases the coefficient accompanying the interaction term is small and not statistically significant. Only after 2009 some event shifted the attendance rates for the treatment group, but clearly not before.

[49] Some additional clarifications must be made in order to assess a fair comparison. Firstly, the average baseline of secondary school attendance in most Latin American countries was considerably lower than that of Argentina. Furthermore, we focus on upper-secondary attendance rates while the evidence presented above corresponds to the whole secondary school level.

Table 5. Probability of attending secondary school

Placebo regressions

	Intervention in				
	2004	**2005**	**2006**	**2007**	**2008**
Treatment*After	***0.0192***	***0.0252***	***0.0179***	***0.0157***	***0.0155***
	(0.0184)	***(0.0167)***	***(0.0140)***	***(0.0128)***	***(0.0151)***
Treatment	*-0.0856****	*-0.0872****	*-0.0796****	*-0.0751****	*-0.0719****
	(0.0166)	*(0.0147)*	*(0.0105)*	*(0.00831)*	*(0.00753)*
After	*-0.0525***	*0.0437*	*0.0489*	*0.0507*	*0.0759**
	(0.0249)	*(0.0392)*	*(0.0393)*	*(0.0386)*	*(0.0421)*
Child and HH head's characteristics	Yes	Yes	Yes	Yes	Yes
Other HH Characteristics	Yes	Yes	Yes	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes	Yes	Yes	Yes
Observations	18,829	18,829	18,829	18,829	18,829

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: OLS estimations. Dependent binary variable: *Attends*, equals 1 if the child attends upper secondary level; *Treatment* equals 1 for eligible children and 0 for non-eligible children; *After* is defined ad-hoc for each year (for example in 2006 it equals 0 in the period 2004 to 2006 and 1 in the period 2007-2009). For a description of control variables included, refer to Table 4.

Alternative definition of the pre-intervention period

As discussed in Section 2, the National Education Law of 2006 extended compulsory schooling for children aged 15-17. Therefore, if this legislation altered schooling incentives differently for the treatment and control groups, then the effects we find cannot be adjudicated solely to the AUH. The results of the placebo experiment with 2006 as the false intervention date – column 3 in Table 5 – represent evidence against this possibility. However, to reinforce the rejection of the incidence of this law we additionally assess the AUH impact on secondary attendance by establishing an alternative shorter pre-intervention period: from 2007 to 2009, rather than 2004 to 2009. Column 1 in Table 6 shows the original results – the same results reported in Table 4, column 3 – while column 2 presents the estimated results when restricting the sample to years 2007-2014 and defining 2007-2009 as the pre-intervention period. Coefficients are quite similar in terms of size

and statistical significance, reinforcing the hypothesis that it was the AUH in 2009 what caused the increase in attendance rates of poor children living in informal households.

Table 6. Probability of attending secondary school

Alternative pre-intervention periods

	Pre-intervention period	
	2004-2009	**2007-2009**
Treatment*After	***0.0388******	***0.0328******
	(0.00885)	***(0.00975)***
Treatment	*-0.0757****	*-0.0707****
	(0.00622)	*(0.00831)*
After	*0.0309*	*-0.100****
	(0.0433)	*(0.0358)*
Child and HH head's characteristics	Yes	Yes
Other HH Characteristics	Yes	Yes
Regional and Time Dummies; Regional Trends	Yes	Yes
Observations	35,002	27,035

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: OLS estimations. Dependent binary variable: Attends, equals 1 if the child attends upper secondary level; *Treatment* equals 1 for eligible children and 0 for non-eligible children; *After* is defined ad-hoc in each model (for column 1 it equals 1 in the period 2010-2014 and 0 for the period 2004-2009, for column 2 it equals 1 for the same period but 0 for 2007-2009). For a description of control variables included, refer to Table 4.

Alternative samples

Since the EPH does not include information to identify AUH beneficiaries we relied on children's eligibility based on their parents' labor status. However, some limitations of the survey may lead to

classification errors. To start with, we do not have information on one or both parents when they do not live within the household[50]. Furthermore, even if parents live with their child it is not always straightforward to identify this relationship given the fact that the EPH collects information on the family linkage of each household member only in terms of the head of household[51].

To assess the extent to which these limitations may affect results, we define three alternative nested samples that account for different possible situations: *(i)* a first sample that only contains those children for whom both parents live in the household; *(ii)* an alternative larger sample that includes children for whom least one parent is present; and finally *(iii)* one that also incorporates those children living in households where neither parent is present. Considering our universe is composed by all children aged 15-17 belonging to the first four income deciles, then sample *(i)* represents 64.4% of that target population, sample *(ii)* adds up a considerable fraction of children leading to a total coverage of 94.1%, while sample *(iii)*, by construction, holds the total universe. In all three samples, whenever more than one adult could be identified as the mother or father of the child, the child was only considered eligible if all of the 'potential' parents met the eligibility conditions[52].

Table 7 shows that the estimated effects of the program are not altered when using these alternative samples, neither in magnitude nor in terms of statistical significance. Given the robustness of the main result to different samples, we choose to conduct the analysis on the basis of sample *(ii)*. Indeed, all the results shown previously relied on this last group of children. The choice is grounded on conceptual reasons. On the one hand, it extends sample *(i)* by including many single-parent households, mostly female headed households, where poverty rates are usually higher and are thus possibly more prone to belong to the treatment group. On the other hand, sample *(ii)* excludes those children for whom we have no information on neither of their parents working conditions – sample *(iii)*. The chosen sample, of course, suffers from the risk of including in the treatment group children that should belong to the control group: when the parent living with the children meets the program's eligibility conditions but the parent not living within the household does not.

[50] The latter generally includes households where grandparents are in charge of their grandchildren.
[51] For instance, suppose a family is composed by the head of household, two of his daughters, two sons in law and a grandson between 15 and 17 years old. In such a case we would not be able to identify who the father and mother of the child are.
[52] In the example set in the previous note, this would imply that both daughters and both sons-in-law should meet the requirements. These cases, however, only represented 0.8% of sample *(i)* and 1.8% of sample *(ii)*.

Nevertheless, even making very pessimistic assumptions, we estimate that only 9% of sample *(ii)* could be wrongly classified in the treatment group[53].

Table 7. Probability of attending secondary school

Alternative samples

	Sample		
	(i)	***(ii)***	***(iii)***
Treatment*After	***0.0358******	***0.0388******	***0.0376******
	(0.00984)	***(0.00885)***	***(0.00865)***
Treatment	*-0.0761****	*-0.0757****	*-0.0792****
	(0.00750)	*(0.00622)*	*(0.00594)*
After	*0.0287*	*0.0309*	*-0.00799*
	(0.0459)	*(0.0433)*	*(0.0134)*
Child and HH head's characteristics	Yes	Yes	Yes
Other HH Characteristics	Yes	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes	Yes
Observations	23,953	35,002	37,207

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: sample *(i)* includes children aged 15-17 for whom both parents live in the household; sample *(ii)* includes children aged 15-17 for whom at least one parent is present; sample *(iii)* includes all children aged 15-17, irrespective of whether both, one or neither parent in present in the household. See Table 4 for a description of the variables included.

6. Heterogeneous effects

Our estimates show that the AUH increased net secondary attendance rates for those eligible children aged 15 to 17 years old by almost 4 percentage points, but heterogeneities may be

[53] This is based on the assumption that all non-present parents live and are recognized as such. Also, we assume that their formality rate is similar to that of parents living with their children – around 36%.

hidden behind this average effect. In this section we explore whether the impact of AUH on attendance rates varies across groups. Firstly, we look for heterogeneous effects by age and gender of children. Secondly, we assess whether the impact is related to household characteristics: number of children and education level of the head of household.

Heterogeneities by age

Table 8 shows that the effect varies considerably across age groups. Compared to the almost 4-percentage- point increase for the group aged 15-17, the effect is only 0.8 for the 12-14 age group – lower secondary. For children aged 6-11 – primary school age – the effect is even smaller but still significant (0.4 percentage points) while for the youths between 18 and 20 years old the estimated effect of the AUH is not statistically significant.

Table 8. Probability of attending school

Heterogeneities by age range

	Age Range			
	6-11	**12-14**	**15-17**	**18-20**
Treatment*After	***0.00422******	***0.00809****	***0.0388******	***0.0170***
	(0.00153)	***(0.00315)***	***(0.00885)***	***(0.0151)***
Treatment	*-0.00383****	*-0.0153****	*-0.0757****	*-0.0867****
	(0.00113)	*(0.00256)*	*(0.00622)*	*(0.00941)*
After	*-0.0229**	*-0.00251*	*0.0309*	*-0.109*
	(0.0120)	*(0.0172)*	*(0.0433)*	*(0.0695)*
Child and HH head's characteristics	Yes	Yes	Yes	Yes
Other HH Characteristics	Yes	Yes	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes	Yes	Yes
Observations	69,332	34,904	35,002	28,792

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: OLS estimations. Dependent binary variable: *Attends*, equals 1 if the child attends the corresponding level; *Treatment* equals 1 for eligible children and 0 for non-eligible children; *After* equals 1 in the period 2010-2014 and 0 for the period 2004-2009. For a description of the variables included, see Table 4.

The latter result is consistent with the fact that individuals older than 18 years old are not eligible for the program, so no effect of the AUH is expected in terms of their schooling. Regarding the age groups covered by the program (6 to 11, 12 to 14 and 15 to 17), the results are consistent with the existing international evidence on the impact of CCT programs on schooling: the effect of the AUH is larger for higher levels of education, where baseline attendance rates are lower (Saavedra and García, 2012; Fiszbein *et al.*, 2009). Indeed, even though the explicit cost of attending school may be similar at all educational levels, the opportunity costs certainly increase with age: older children may work in the labor market or allow for other adults in the household to do so by taking care of younger siblings or performing other household chores[54]. Therefore, it is plausible that the economic incentives introduced by the AUH may have lower or even insignificant effects for younger school-aged children whose educational decisions are less sensitive to economic changes, thus explaining the larger impact for the oldest eligible children.

We also explore whether there are age heterogeneities among 15-17 year olds. Table 9 shows the results when zooming into this group. We find an unclear pattern: the impact of the AUH seems to be concentrated in those aged 15 and 17.

[54] The legal minimum working age in Argentina is 16 years old (*Ley de empleo infantil 26,390*).

Table 9. Probability of attending secondary school

Heterogeneities by age

	Age		
	15 years old	***16 years old***	***17 years old***
Treatment*After	***0.0425******	***0.0252***	***0.0518******
	(0.0132)	***(0.0165)***	***(0.0162)***
Treatment	*-0.0718****	*-0.0537****	*-0.103****
	(0.00942)	*(0.0100)*	*(0.00944)*
After	*-0.0286*	*-0.0442*	*0.0593*
	(0.0372)	*(0.0490)*	*(0.0506)*
Child and HH head's characteristics	Yes	Yes	Yes
Other HH Characteristics	Yes	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes	Yes
Observations	12,481	11,354	11,167

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: for a description of the variables included, see Table 4.

Heterogeneities by gender

Table 10 shows that the increase in attendance rates was mostly driven by improvements in boys' attendance: the estimated impact for boys is well above 5 percentage points while that of girls is below 2 percentage points and not statistically significant.

Once again, more than one mechanism may explain these results. As stated before, different baseline levels of attendance may be in part responsible. In fact, initial attendance rates were lower for boys: around 70% as compared to 80% for girls among the treatment group. Also, according to the literature, family decisions on girls' schooling seem to be more tied to cultural factors which are less affected – at least in the short term – by changes in household income. For instance, previous evidence for Argentina (Sosa Escudero and Marchionni, 1999) suggests that girls' attendance is rather inelastic as compared to boys'.

Table 10. Probability of attending secondary school

Heterogeneities by gender

	Boys	Girls
Treatment*After	***0.0583****	***0.0165***
	(0.0108)	***(0.0122)***
Treatment	*-0.100****	*-0.0499****
	(0.00813)	*(0.00731)*
After	*0.0263*	*0.00823*
	(0.0393)	*(0.0696)*
Child and HH head's characteristics	Yes	Yes
Other HH Characteristics	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes
Observations	17,822	17,180

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Note: for a description of the variables included, see Table 4.

Heterogeneities by household characteristics

Table 11 shows the AUH effect on net school attendance among families which differ in the number of children. Despite the fact that the impact is statistically significant for all groups, it increases with the number of children. In particular, the effect for larger households (5 or more children) almost doubles that of families with one or two children. This result is consistent with the fact that more eligible children in the household imply higher benefits and a potentially larger income effect of the AUH. Thus, larger families may show more commitment with the conditionalities of the program.

Table 11. Probability of attending secondary school

Heterogeneities by number of children in the household

	Number of Children		
	1 or 2	***3 or 4***	***5 or more***
Treatment*After	***0.0267****	***0.0326****	***0.0506****
	(0.0130)	***(0.0143)***	***(0.0218)***
Treatment	*-0.0612****	*-0.0686****	*-0.104****
	(0.0106)	*(0.00872)*	*(0.0127)*
After	*-0.0760**	*0.0652*	*0.0535*
	(0.0390)	*(0.0451)*	*(0.0700)*
Child and HH head's characteristics	Yes	Yes	Yes
Other HH Characteristics	Yes	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes	Yes
Observations	13,799	14,301	6,902

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Note: for a description of the variables included, see Table 4.

Finally, table 12 explores whether the effect varies with household structure –two-parent or single-parent families– and with the education level of the head of household. We find a positive impact of AUH on school attendance for all groups, and the effect appears to be slightly larger among children with low educated head of household.

Table 12. Probability of attending secondary school

Heterogeneities by characteristics of the head of household

	Single-Parent	Two-Parent	Low Education	High Education
Treatment*After	***0.0369****	***0.0398*****	***0.0360*****	***0.0240***
	(0.0157)	***(0.0105)***	***(0.0104)***	***(0.0142)***
Treatment	*-0.0685****	*-0.0765****	*-0.0841****	*-0.0435****
	(0.00970)	*(0.00758)*	*(0.00752)*	*(0.0100)*
After	*-0.00686*	*-0.00582*	*0.121***	*0.0269**
	(0.0300)	*(0.0128)*	*(0.0549)*	*(0.0153)*
Child and HH head's characteristics	Yes	Yes	Yes	Yes
Other HH Characteristics	Yes	Yes	Yes	Yes
Regional and Time Dummies, Regional Trends	Yes	Yes	Yes	Yes
Observations	10,994	24,008	25,505	9,497

*Clustered robust standard errors in parenthesis; * p<0.10, ** p<0.05, *** p<0.01.*

Source: own estimations based on *Encuesta Permanente de Hogares*.

Notes: OLS estimations. "Low Education" includes household which head has less than secondary school education, "High Education" refers to households where the head completed secondary education. For a description of the variables included, see Table 4.

7. Concluding remarks and further research

Argentina has traditionally stood out in terms of educational outcomes among its Latin American counterparts. Schooling of older children, however, still shows room for improvement especially among the more vulnerable school-age children. Fortunately, during the last years a sizeable improvement in attendance rates for children aged 15 through 17 took place. This could be related to the 2006 National Education Law that made upper-secondary education compulsory. In this paper, instead, we show that the *Asignación Universal por Hijo* (Universal Child Allowance, AUH), a massive conditional cash transfer program implemented in 2009 in Argentina, may be mostly responsible for this improvement.

Using a difference-in-difference strategy based on data from the Argentinean National Permanent Household Survey we estimate that the program accounts for a 3.9 percentage point increase in secondary school attendance among eligible children aged 15 through 17. This effect is robust to different specifications and a large set of checks. Also, we present evidence suggesting that this effect is not related to the expansion of compulsory education that took place in Argentina in 2006. Moreover, the positive impact of the AUH in attendance rates is not homogenous: the effect seems to be driven particularly by boys and is higher for children living in larger households where the head has low educational attainment.

Further research should point in several directions. A first relevant issue would be to unravel which mechanisms within the AUH are responsible for the increase in attendance rates. The effect may be driven by the monthly benefit itself or by the conditionality, or both mechanisms could be operating simultaneously. A deep understanding of these alternative channels is indeed relevant in terms of improving the design of CCT programs. Secondly, it would be interesting to explore if the AUH has not only increased secondary school attendance among eligible children but also affected other educational results, such as intra-annual dropouts or secondary school completion rates. Thirdly, it would also be relevant to disentangle if this increase in attendance rates is matched by a similar result in the employment realm. It could be expected that an increase in attendance rates may contribute to a reduction in labor participation among the 15-17 age group. It could also be the case, however, that those upper-secondary school aged children were not working in the labor market before the AUH, but in charge of household chores such as taking care of their siblings. In that case the AUH may be altering instead other members' labor participation. Although household decision processes are certainly difficult to assess, exploring these hypothesis would shed light on the mechanisms that are at work and thus further refine the AUH's design.

References

Acemoglu, D. and Angrist, J.: How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws. NBER Macroeconomics Annual, *15*,.9-59 (2000)

Agencia Nacional de Seguridad Social (ANSES): La Asignación Universal por Hijo para la Protección Social. Boletín Cuatrimestral. I Cuatrimestre 2014. Observatorio de la Seguridad Social. (2014)

Alzúa, M.L., Gasparini, L. and Haimovich, F.: Educational Reform and Labor Market Outcomes: the Case of Argentina's Ley Federal de Educación. Journal of Applied Economics, 18(1), 21-46 (2015)

Angrist, J. and Krueger, A.: Does Compulsory School Attendance Affect Schooling and Earnings? Quarterly Journal of Economic, 106(4), 979-1014 (1991)

Angrist, J. and Pischke, J.: *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton University Press (2009)

Auguste, S.: La calidad educativa en la Argentina. Documento de Trabajo 116, Fundación de Investigaciones Económicas Latinoamericanas, Buenos Aires (2012)

Attanasio, O., Fitzsimons, E., Gómez, A. and López, D.: Child education and work choices in the presence of a conditional cash transfer programme in rural Colombia. (WP06/13) London, Institute for Fiscal Studies (2008)

Bastagli, F.: Conditionality in public policy targeted to the poor: promoting resilience? *Social Policy & Society, 8(1)* (2008)

Becker, G. S.: *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* 2nd Edition, The National Bureau of Economic Research (1975)

Boffi, S.: Políticas de transferencias de ingresos. Su impacto en el mercado de trabajo. Experiencias en Latinoamérica y los casos de la Asignación Universal por Hijo y el Programa de Inclusión Previsional en Argentina. Publicación No 16, Centro de Investigación en Trabajo, Distribución y Sociedad (2013)

Cecchini, S.: Educación, Programas de Transferencias Condicionadas y Protección Social en América Latina y el Caribe. In Feijóo, M. and Poggi, M. (eds.) *Educación y Políticas Sociales. Sinergias para la Inclusión*. IIPE-UNESCO (2014)

Cigliutti, M., Echeverri Gómez, M., Golinsky, F., Gutiérrez, A. and Sorá, M.G.: Conditional Cash Transfer Programs: Measuring the Impact on Education. Anales de la Asociación Argentina de Economía Política, L Reunión Anual, Salta, Argentina (2015)

De Brauw, A. and Hoddinott, J.: Must conditional cash transfer programs be conditioned to be effective? The impact of conditioning transfers on school enrollment in Mexico. IFPRI, Discussion Paper, 00757 (2008)

D'Elia, V. and Navarro, A.I.: Universal Child Allowance and School Delay of Children in Argentina. *Revista de Análisis Económico*, *28(2)*, 63-90 (2013)

Dirección Nacional de Información y Evaluación de Calidad Educativa (DiNIECE), http://portales.educacion.gov.ar/diniece/

DiNIECE: La Transformación del nivel secundario (2006–2009). Temas de Educación. Boletín DiNIECE year 6, no. 9, Dirección Nacional de Información y Evaluación de la Calidad Educativa, Ministerio de Educación, Buenos Aires, Argentina (2011)

Eckstein, Z. and Zilcha, I.: The effects of compulsory schooling on growth, income distribution and welfare. Journal of Public Economics, 54(3), 339-359 (1994)

Escobar, A. and De la Rocha, M.: Girls, mothers and poverty reduction in Mexico: evaluating Progresa-Oportunidades, en S. Razavi (ed.), *The Gendered Impacts of Liberalization: towards embedded liberalism?* Nueva York, Routledge (2008)

Fiszbein, A., Schady, N., Ferreira, F.H.G., Grosh, M., Keleher, N., Olinto P. and Skoufias, E.: *Conditional Cash Transfers: Reducing Present and Future Poverty*. World Bank Publications, The World Bank, number 2597 (2009)

Gasparini, L. and Cruces, G.:. Pobreza y Políticas de Transferencias Monetarias en Argentina. Diagnóstico y alternativas. Mimeo, CEDLAS (2015)

Goldin, C. and Katz, L.: Mass Secondary Schooling and the State: The Role of State Compusion in the Hight School Movement. NBER Working Paper 10075 (2008)

Lleras-Muney, A.: The Relationship Between Education and Adult Mortality in the United States. Review of Economic Studies *72*, 189-221 (2005)

Marchionni, M. and Alejo, J.: Funding Basic Education Throughout the Demographic Transition in Argentina (Chapter 8). In *As time goes by in Argentina. Opportunities and Challenges of the Demographic Transition*. Gragnolati, M., Rofman, R., Apella, I. y Troiano, S. (eds.). World Bank Group, Washington DC, pp. 203-234 (2015)

Oreopoulos, P.: The compelling effects of compulsory schooling: evidence from Canada. Canadian Journal of Economics, Canadian Economics Association, 39(1), 22-52 (2006a).

Oreopoulos, P.: Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter. The American Economic Review, *96 (1)*, 152-175 (2006b)

Parker, S.: Evaluación del impacto de Oportunidades sobre la inscripción escolar: primaria, secundaria y media superior. *Resultados de la Evaluación Externa del Programa de Desarrollo Humano Oportunidades 2002*, Mexico D.F., Secretaría de Desarrollo Social (SEDESOL) (2003)

Pautassi, L., Arcidiácono, P. and Straschnoy, M.: Asignación Universal por Hijo para la Protección Social de la Argentina. Entre la satisfacción de necesidades y el reconocimiento de derechos, CEPAL, Serie Políticas Sociales N°184 (2013)

Paz, J. and Golovanevsky, L.: Programa Asignación Universal por Hijo para la Protección Social. Un ejercicio de evaluación de sus efectos sobre los hogares. III Jornadas Nacionales sobre Estudios Regionales y Mercados de Trabajo, Jujuy, Argentina (2014)

Ruiz, M.C. and Schoo, S.: La obligatoriedad de la educación secundaria en América Latina. Convergencias y Divergencias en cinco países. Foro de Educación, 12(16), 71-98 (2014)

Saavedra, J.E. and García, S.: Impacts of Conditional Cash Transfer Programs on Educational Outcomes in Developing Countries: A Meta-analysis. Rand Working Paper Series (2012)

Secretaría de Desarrollo Social (SEDESOL): *Oportunidades, un programa de resultados*. México, D.F. (2008)

Socio-Economic Database for Latin America and the Caribbean (SEDLAC). Centro de Estudios Distributivos Laborales y Sociales (CEDLAS) and the World Bank, http://sedlac.econo.unlp.edu.ar/eng/

Sosa Escudero, W. and Marchionni, M.: Household Structure, Gender and the Economic Determinants of School Attendance in Argentina. Anales de la Asociación Argentina de Economía Política, XXXIV Reunión Anual, Rosario (1999)

UNICEF: Acerca de la Obligatoriedad de la Escuela Secundaria en Argentina. Análisis de la Política Nacional (2012)